

Not just academic

The Russian Academy of Sciences is failing to provide either the quality of research or the scientific advice that reformers had been hoping for.

Thirteen years after the collapse of the Soviet Union, the Russian Academy of Sciences has so far been unable to reform itself into an organization that can provide the leadership that the nation's scientific community needs.

The academy's recent advice to Russian president Vladimir Putin that the Kyoto Protocol on climate change had "no scientific basis" (see page 12) is just one example of the academy's weakness. The advice didn't reflect the views of climate scientists inside or outside Russia. It is unlikely to have much effect on Putin's final decision on whether to ratify the Kyoto agreement, but it reflects badly on an academy whose centralized decision-making still has a Soviet flavour.

The Russian academy has two main functions, each vital to the country's development: it runs some 450 institutes, employing thousands of scientists, and is supposed to be the main body providing impartial scientific advice to the Russian government.

The academy traditionally comprised the élite of Russia's scientific intelligentsia. But in common with other public-sector enterprises in Russia, it has suffered a dramatic decline in wealth and reputation following the end of communism and the collapse of the Soviet military-industrial complex. Its institutes receive more than half of Russia's meagre public support for research and the great bulk of all money available for basic science. But critics say that it provides a poor return on this investment. Many of its institutes are isolated from Russia's real needs in industry or health, for example. They offer little in terms of technology innovation and are barely involved at all in science education.

A number of the academy's scientists still do good research in fields such as mathematics and some branches of physics, but the

academy no longer produces a wide range of internationally competitive science. The most pressing priority for the heads of its institutes is to maintain their physical assets and political influence.

The academy is vital to the future of basic research in Russia. In a country where university-based researchers have little prospect of receiving stable research funds, the academy can still provide a home for the next generation of scientists. In order to do this effectively, it needs to reach out and strengthen its links with universities, industry, government agencies and hospitals. But there are far too many institutes and not enough resources to run them. Reform has been regarded as politically impossible in the past, but is now perilously overdue. An outside review of every institute's research performance, drawing on international expertise, could help determine which institutes are worth investing in.

It is clearly in Russia's interest to allow scientists to work in up-to-date laboratories where academic freedom is respected. The country badly needs strong, self-assured and unbiased institutions, and a renewed and streamlined academy could become one of them. The atmosphere in Russia is not conducive to the development of such institutions, however. The July murder of journalist Paul Klebnikov, editor of the Russian edition of *Forbes* magazine, after it had published a list of the country's 100 wealthiest individuals, served to reinforce the perception that the political climate in Russia is growing ever more oppressive.

In this context, it is hard to see a way forward for Russian science. Academy researchers and managers must continue to strive as best they can for a better environment, in which their work will be properly supported and their advice garnered fairly, and taken seriously. ■

Distributing the costs of climate change

Policy-makers must face up to the fact that global warming is creating winners and losers.

As with any profound disruption of society, such as war or hyperinflation, the changes brought about by our exploitation of fossil fuels will produce winners and losers. A session on climate extremes and their impacts, organized by *Nature* at the EuroScience Open Forum in Stockholm last week (see *Nature* 430, 277; 2004), discussed who will, and who should, pay for damage caused by global warming.

There are two basic strategies for dealing with global warming: climate-change mitigation, and adaptation to the changing conditions. For the next few decades, we will have no choice but to adapt. According to the 2001 report of the Intergovernmental Panel on Climate Change (IPCC), atmospheric carbon dioxide levels will rise until 2030 whether future emissions are curbed or not; the greenhouse gases that have already been emitted will dominate our medium-term future. But mitigation now can make a significant difference to the adaptation costs that will face our grandchildren.

So who benefits and who will have to pay? The question must be considered for each of the various effects of climate change. Take, for example, an increase in intense precipitation events, which is likely to

occur in many areas over the next century, according to the IPCC assessment. Assuming that the main damage of intense rains lies in the flooding of land, those who live in a floodplain are most likely to lose money — in many European countries, flood damage is not insured. In most rich nations, people can choose between the pleasures of the riverside and the safety of a hill-top, so one might argue that floodplain dwellers have only themselves to blame. But a country such as the Netherlands does not offer many flood-proof building sites and, outside affluent societies, individual choices are limited.

There are winners too, including the builders who repair the damage and oil companies that benefit indirectly through the unrestricted sale of fossil fuels. But anyone who uses energy from fossil fuels at a price that does not account for climate-related costs of greenhouse-gas emissions is also 'winning' at someone else's expense. Winners and losers may be the same people, but usually they are not.

Inevitably, it is up to policy-makers to ensure that the costs of mitigation and adaptation are at least partly borne by those to whom climate change is attributable. Whatever the future of international treaties, lawyers and climatologists have interesting times ahead. ■





Open door
European meeting
gets the Swede smell
of success
p5



Kill or cure?
Feathers fly in
Thailand over
bird flu vaccine
p6



Chipping in
Semiconductor
industry set to survey
health of workers
p7



Room to roam
Reserve may offer
pandas a chance to
meet new friends
p9

Ethics review slams government panels over conflicts of interest

Emma Marris, Washington

A US government ethics office is demanding that agencies comply more fully with rules requiring experts on advisory panels to declare potential conflicts of interest.

Some two-thirds of scientists, engineers, lawyers and other specialists on the panels don't fill out any financial conflict-of-interest paperwork. For years, government agencies have chosen to give them a special job title — 'representative' — that exempts them from declaring any conflicts of interest they may have.

The loophole affects hundreds of panels that advise the government on everything from lead concentrations in drinking water to the safe disposal of nuclear weapons.

Most of the panels were set up under the Federal Advisory Committee Act (FACA) — a pioneering 1972 law passed to ensure that advice is given to the government publicly and transparently. But a congressional audit earlier this year found that agencies have come to flaunt the spirit of the law by using a bureaucratic dodge that prevents many possible conflicts from being declared.

FACA panels on contentious issues such as climate change, earthquake risk assessment or food safety are watched avidly by interest groups and the press. But the audit said that committees offering advice on these very topics frequently consisted entirely of representatives who didn't have to declare any conflicts of interest.

"The system has got completely out of whack over the years," says Merrill Goozner, an analyst at the Washington-based Center for Science in the Public Interest, a watchdog group specializing in dietary issues. He says he hopes that the memo from the Office of Government Ethics, which was released in July as a follow-up to the audit published by the Government Accountability Office (GAO) in April, will spur reform. "Everyone is now on notice," he says.

Jennifer Sass, a spokeswoman for the Natural Resources Defense Council, an environmental group, says that everyone giving scientific advice should be asked about conflict of interest, and that those with a clear bias should be denied voting rights on the panels.



A representative sample? Critics are worried about the financial interests of US advisory panel members.

The congressional audit was requested by two Democratic members of Congress who were following up on reports that the Bush administration was filling its advisory committees with political allies. But the audit instead highlighted the practice of appointing many panel members as representatives — a practice it described as "long-standing and firmly rooted in agency culture".

Inherent bias

Representatives are supposed to stand for an industry or interest group — a panel on food safety, for example, might have a representative from the beef industry. Because their bias is obvious and expected, they don't fill out conflict-of-interest paperwork. But many agencies overuse the designation, according to the GAO, and the departments of the interior, energy and agriculture appoint all panel members as representatives, including people specifically chosen to provide unbiased scientific advice.

The ethics-office memo is already having some effect, according to agency officials. Shayla Simmons, ethics official at the Department of the Interior, says her depart-

ment is now reviewing each new committee member. She says that many panel members just aren't expected to give objective advice. "We will still have the majority of the people on committees serving as representatives," she says, "because most people are representing some interest."

Julie Quick, a spokeswoman from the Department of Agriculture, says that change is in the air there also. "We are considering a model used by other agencies," she says, in which academic scientists can no longer be appointed as representatives. A spokesperson for the energy department, which oversees panels dealing with everything from nuclear clean-up to research priorities in physics, would say only that the issue was "under review".

In a report due out this November on presidential advice, the National Academy of Sciences is expected to take up the question of how conflicts of interest are declared. The academy's multi-stage process for assessing conflicts on its own panels has been cited by critics of the federal system as a possible model, although others say the procedures are inconsistently applied in practice. ■

Disaster movie highlights transatlantic divide

Quirin Schiermeier, Munich

The world is in crisis. Hurricanes rage and temperatures are plunging fast. Icy-cold water has swamped the streets of New York. This cinematic portrayal of the next ice age may sound harrowing, but it has left some movie-goers floundering.

In a twist worthy of a Hitchcock plot, the disaster movie *The Day After Tomorrow* made audiences in Germany less worried about the effects of global warming and climate change. In contrast, audiences in the United States have had their fears fuelled by the film, hinting at widely different perceptions of climate change either side of the Atlantic, according to two unpublished studies.

Since its release in May, more than 30 million adults have seen *The Day After Tomorrow*. The film portrays the rapid arrival of an ice age after climate change shuts down the existing circulation pattern in the North Atlantic.

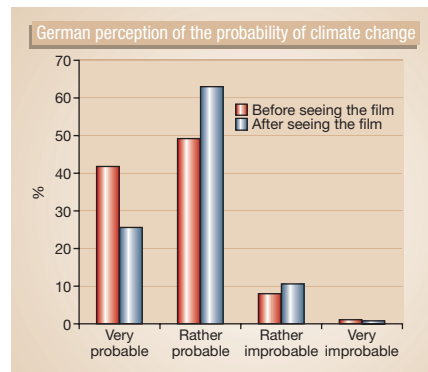
Currently the fifth most successful film of the year — it has grossed some US\$540 million worldwide — *The Day After Tomorrow* was billed as science fiction. Despite this, its plot bears a faint resemblance to what climate researchers think might happen in a future greenhouse world (see *Nature* 429, 347–348; 2004). But allegations of half-baked science and panic-mongering in the movie have sparked heated debate.

Fritz Reusswig, a sociologist at the Potsdam Institute for Climate Impact Research, asked 1,200 movie-goers throughout Germany to fill in questionnaires immediately before and after they went to see the film. Most of the interviewees said that they believed climate change to be real, but the level of conviction dropped significantly after they had seen the movie (see graph).

"In a sense, the film has been counter-productive," says Reusswig, adding that the seemingly unlikely connection between 'global warming' and 'cold' may have seeded



Washout: German audiences were not convinced by unrealistic events in *The Day After Tomorrow*.



disbelief among the audience. "Most people here associate climate change with heatwaves and floods," he says. "The film has made them ask: 'If this is what climate change is like, then we are no longer sure it is real'."

But the US study — funded by the National Science Foundation — shows a different picture. Anthony Leiserowitz of Decision Research, a risk-research company based in Eugene, Oregon, conducted two

nationwide surveys — a week before and four weeks after the movie's release — to explore its public impact.

The results, which will be published in the November issue of *Environment*, revealed no evidence for a significant shift in overall public opinion, despite the media hype for the film. But of the 529 people interviewed after the film's release, half of the hundred or so who had seen it said that it had made them "somewhat or much more worried" about global warming, whereas only 1% said that they became less worried. The film even had a statistically significant impact on voting intentions among the film viewers — the survey found that overall they became slightly more likely to vote for Democrat presidential candidate John Kerry this November.

"Without doubt *The Day After Tomorrow* has raised awareness around the issue of climate change," says Katie Mandes, a spokeswoman for the Pew Center on Global Climate Change in Arlington, Virginia, a global-warming think-tank. The centre's website, which explains the scientific scenario the movie depicts, more than doubled its traffic after the film's release, she says.

The different perceptions in Europe and the United States may reflect the respective levels of awareness, Mandes suggests. "Europeans are light-years ahead in accepting that climate change is a real phenomenon," she says. So most Europeans probably had a fixed notion of the effects of climate change before they went to see the film — meaning that the movie's version of events was likely to contradict their expectations.

Those waiting to watch *The Day After Tomorrow* at home will be treated to an extra slice of science education: the DVD, scheduled for release in October, will feature a bonus climate documentary by Los Angeles film-maker Charles Kiselyak.

Newspapers fog public view of climate change

The US public's perception of global warming is skewed by something more pervasive than the occasional Hollywood movie, says a study released late last month. Its authors argue that people have been misled by newspaper reports that tend to give equal weight to both sides of the climate-change debate.

In the study, Maxwell Boykoff, a graduate student in environmental studies at the University of California, Santa Cruz, and his brother Jules, a visiting assistant professor of politics at Whitman College, Washington, looked at more than 600 randomly chosen articles about climate change from four major newspapers written between 1988 and 2002.

They found that 52.7% of the articles gave "roughly equal attention" to views that humans contribute to global warming and that climate change is just a result of natural fluctuations. Only 35.3% of the articles emphasized the role of humans while presenting both sides of the debate (M. T. Boykoff and J. M. Boykoff *Global Environ. Change* 14, 125–136; 2004).

Although it is standard journalistic practice to give both sides of a story, the Boykoffs say that this has warped public perception of an issue in which there is a strong scientific consensus. "In effect, the press has provided 'balanced' coverage of a very unbalanced issue," says Maxwell Boykoff.

Nicola Jones



Café society: the EuroScience Open Forum won participants over with its relaxed atmosphere.

Organizers claim success for Stockholm science jamboree

Alison Abbott, Stockholm

Sceptics said it couldn't work. The very time and place of the first EuroScience meeting bringing together scientists, journalists and the public to celebrate science — à la the annual AAAS meeting in the United States — spoke against it.

It is difficult enough to get European researchers, who will happily attend US meetings, to come together on home turf. And on a continent where August holidays are sacred, who would instead visit one of Europe's most expensive cities for an untested meeting?

But 1,800 people came to the EuroScience Open Forum (ESOF) meeting in Stockholm, Sweden, from 25–28 August, exceeding the 1,500 expected. And they soon found their irritation at some teething problems waning in the feel-good atmosphere that ESOF generated — aided by its location at the heart of the historic city. "Being in the city is fundamental for this type of conference," says Clive Cookson, science editor of the London-based *Financial Times*.

ESOF 2004 was organized by EuroScience, founded in 1997 to promote science and its public discussion. The meeting's sponsors, including *Nature*, rather than its participants, met most costs. Sponsors expressed satisfaction, and the two German groups who contributed most, the Robert Bosch Foundation and the Stifterverband, a donors' association for the sciences and humanities, are supporting the next meeting, planned for Munich in July 2006.

A third of participants came from Sweden and more than 50 flew in from the United States, with 68 countries represented in all. There were 350 journalists.

Alan Leshner, chief executive of the AAAS, formerly the American Association for the Advancement of Science, took the opportunity to announce that the body will in future use only its acronym, to emphasize its global reach.

Though many attendees complained of too many sessions on media and policy and too few on science, some of the scientific sessions were novel and successful. Rainer Goebel from the University of Maastricht in the Netherlands, for example, made his audience gasp when he described his studies of people controlling a computer game with their thoughts via brain-imaging machines.

Another success was the Science in the City programme, which brought theatre, films, exhibitions and other events to venues across Stockholm. The Amazing Profmobile, a rickshaw, took scientists to address surprised shoppers and pedestrians. Rolf Tarrach, a physicist from the University of Barcelona, rushed gleefully between theatre and scientific sessions. The programme was a "fantastic European experience which resonates with this historic city", he says.

There were a few teething troubles. Journalists complained that talks lacked written summaries. Some of the presentations were too technical for their audience. And few students and younger researchers were present.

There will be better support next time for journalists, and more training of speakers, says ESOF 2004 organizer Carl Sundberg, a physiologist at the Karolinska Institute in Stockholm who is also on the steering committee for ESOF 2006. "And we'll look more imaginatively at participation fees, to ensure there is no barrier to students," he says. ■

Wildlife campaigners fight planned site for Scripps Florida

Rex Dalton, San Diego

An environmental backlash is threatening to delay the Scripps Research Institute's plans to build a huge research and biotechnology complex near Palm Beach in Florida.

Scripps Florida is to be built on a 780-hectare site near a wildlife refuge. But criticism from environmental groups has made some local politicians change their minds about the site. State and county officials plan to invest \$770 million in Scripps Florida (see *Nature* 426, 4; 2003).

Palm Beach County last month asked Scripps, based in La Jolla, California, to consider two other sites, both of which are less environmentally sensitive. But the sites are smaller, at less than 350 hectares, and might curtail the institute's industrial growth.

Scripps' board will meet privately on 13 September to consider its options. Although it could agree to one of the new sites, under the present contract signed by the county, it can demand permits for the original property by next January. Palm Beach County council will meet on 14 September to consider its response to Scripps' decision.

If Scripps' board pushes for the original site, it faces an environmental fight in the courts, which critics of the project say could hold it up for years.

It is also possible that such a course would embolden local opposition to plans laid by Palm Beach County officials, who say they intend to spend \$400 million on Scripps Florida. State governor Jeb Bush, the president's brother, has championed Florida's \$370-million state contribution to the project.

Already, several environmental groups, including the Florida Wildlife Federation and the National Audubon Society, are challenging the original site's water permits, which could delay the process several months. Called Mecca Farms, the site is next to the J. W. Corbett Wildlife Refuge, in a region crossed by waterways that feed the Everglades.

Chemist Richard Lerner, president of Scripps and mastermind of the Florida project, says that he won't make a site recommendation to his board. "There are pros and cons to every side," he says. Scripps Florida is already hiring scientists who are setting up labs in temporary facilities in Palm Beach County. "The scientists don't care much about which site is selected," Lerner says. ■

Thailand faces dilemma over bird flu vaccine

David Cyranoski, Tokyo

As Thailand's battle with avian influenza drags on, the government has started to crack down on the illegal use of bird vaccines, with several high-profile arrests last month. But farmers are thought to be relying on black-market vaccines in a desperate attempt to avoid culling their chickens, leaving the country debating whether vaccination should be allowed.

For Thailand, once one of the world's major poultry exporters, the outlook is bleak. More than 300,000 chickens have been destroyed since July, and the bird flu that began this summer remains active in 25 villages.

To farmers trying to protect their livelihoods, vaccination seems like a good idea. It promises to keep healthy birds infection-free and to lessen symptoms in those already infected.

But the unregulated use of vaccines can do harm, says Wantanee Kalpravidh, consultant on avian flu for the United Nations Food and Agriculture Organization in Bangkok. Vaccinated birds that seem healthy can still get infected and spread the virus. And in backyard farms, where chickens run around freely in close proximity to people and other animals, this can spell disaster, she says. "Farmers might not be aware of the risk. It's difficult to keep track of what's happening," she warns.

Black-market vaccines might also contain viruses that have not been properly inactivated, adds Ilaria Capua, who heads a World



The Thai government insists that sick birds are culled.

Organization for Animal Health (OIE) laboratory in Padua, Italy. "You have no idea what you are buying," she says. The vaccine virus strain might also interact with any infectious flu strains present in the same bird, possibly leading to the evolution of a more dangerous virus. A long-term study of avian vaccines in Mexico has recently shown evidence of this

(C.-W. Lee, D. A. Senne and D. L. Suarez *J. Virol.* **78**, 8372–8381; 2004).

Culling is more straightforward and has proved effective — in developed countries at least. When a Texas chicken farm had outbreaks of avian flu in May and June this year, the area was isolated and all the chickens killed. By 25 August the United States was in the clear.

Experts note that vaccines can stop the spread of disease — if accompanied by strict monitoring and testing. This has been successful in parts of Italy and the United States. But many doubt that Thailand has the necessary resources.

Another strike against vaccines is economic, says Alex Thiermann, president of the OIE's Animal Health Code Commission in Paris. As long as a country is vaccinating, it will not get approval for export. This should change in May 2005, says Thiermann, when the rules are expected to become more flexible for developing countries such as Thailand, allowing them to continue trading when vaccinating safely.

Thailand is now deciding what to do. A technical committee reporting to the deputy prime minister is weighing the evidence for and against vaccination, says Kalpravidh. It is expected to reach a decision in the next few weeks.

In the meantime, police have raided market stalls that distribute vaccines to farmers, handing out heavy fines. An education campaign is also planned to explain the risks and benefits of vaccines. But experts doubt that this will stop the use of these vaccines entirely. ■

Lab chiefs fear European rules will cost postdoc jobs

David Osumi-Sutherland, London

British researchers are expressing concern that European Union (EU) rules intended to give workers more job security could wreck the careers of postdocs in their labs.

The EU directive on fixed-term contracts became part of UK law in October 2002. According to Britain's research councils, it means that a postdoc must be offered a permanent contract after four years. In addition, lab chiefs must write a justification for not offering such a contract after three years. The first of these reviews will happen in October 2005.

But as lab chiefs face up to these requirements, they are complaining that they won't be able to increase the number of permanent contracts. The rules will force them to dismiss good postdocs, they say.

"If you are training group leaders they

need to have publications. Three years is totally unrealistic, four is on the edge," says Alan Hall, head of the Medical Research Council's Laboratory for Molecular Cell Biology at University College London. "If it had been five then it might not have caused so many problems."

Mark Marsh, who runs an HIV lab at the MRC facility, says that he is worried about the effects on recruitment. "We have to tell potential postdocs that we can't guarantee more than 3–4 years," he says. "What if it takes 4½ years to get that crucial paper?"

The UK postdoc system — like that of the United States — has traditionally been very flexible. Contracts lasted as long as funding was available and the supervising researcher was happy with the work. But the system has been criticized for trapping postdocs in a series of short-term contracts,

with no job security and little prospect of obtaining a permanent position.

Hall defends the old system. "The success of my research group depends on new young people coming to the group with new ideas and new directions," he says. He draws a distinct contrast with the situation in France, which he sees as a rigid system that depends mainly on permanent research staff.

The EU rules aren't specific to research, but were introduced as part of a general effort to push employers away from a reliance on short-term contracts. No one at the research commission in Brussels was available to comment on the rules' implications for science. But some unions that represent postdocs have argued for steps that would outlaw their indefinite retention on short-term contracts. ■



Making the count: a US study is to explore the cancer rates of workers in semiconductor factories.

Microchip industry proposes broad survey of worker health

Geoff Brumfiel, Washington

Makers of microchips are planning a study of more than 200,000 people to determine whether workers in semiconductor plants have an increased risk of cancer.

The Semiconductor Industry Association, a trade group representing nearly 100 US-based companies, announced on 20 August that it was seeking proposals for the study, which will use the health records of workers in US microchip plants.

The study is planned partially in response to hundreds of recent lawsuits brought by former employees, who allege that work in semiconductor factories gave them cancer. According to John Greenagel, the association's director of communications, the goal is to conclusively determine whether factory workers are more likely to get cancer. "We want the facts," he says.

But some researchers question whether the study can be truly impartial in the hands of an industry trade group that is representing defendants in the ongoing lawsuits. "Anything published will be highly screened by the industry sponsors," claims Joseph LaDou, who specializes in occupational medicine at the University of California, San Francisco.

Materials that can cause cancer if ingested, such as arsenic and beryllium, have been used for decades in semiconductor manufacturing, but until now only a handful of studies have looked for increased cancer risks among employees. In 2001, a study by Britain's Health and Safety Executive of more than 4,000

workers at a National Semiconductor plant in Greenock, Scotland, found that female workers were two to three times more likely than other women in the area to have lung cancer.

A more recent study of deaths among roughly 30,000 former IBM employees in the United States showed an increased incidence of cancer. But that study was pulled from an Elsevier journal earlier this summer before it could be published (see *Nature* **429**, 687; 2004).

"We hope this new study will either give us something to work on or provide reassurances to people that the industry is safe," says Greenagel. He says that the research will be conducted independently. "They will have rights to publish their findings," he says of the researchers. But he adds that the scientists will not be able to identify specific companies in their research. "There will be confidentiality restrictions," he says.

For his part, LaDou remains doubtful and says he wishes the study was being conducted independently by a government agency. "If this had been placed in the hands of the National Institute for Occupational Safety and Health or the National Cancer Institute, we would all be relieved to see it move ahead," he says.

But other researchers are cautiously optimistic. "I think it's a positive step," says Richard Clapp, an environmental health expert at Boston University, who led the unpublished IBM study. "If you can get good, independent scientists, I think it's possible to get honest results."

Five-year grant gets bird database off to a flying start

Erika Check, Washington

Snakes, lizards and bears all have comprehensive specimen databases — and now one is being set up for the bird family.

On 1 August, the National Science Foundation gave a consortium of North American museums a five-year, \$1.5-million grant to build ORNIS — the Ornithological Information System. ORNIS will be a database that contains information on around 4 million bird specimens from 30 collections. Its creators say they will improve on other databases by writing software to automate tasks such as synchronizing and checking data.

Species databases use a method called georeferencing to allow researchers to compare the geographical origins of specimens across collections. In databases such as MaNIS, which is for mammals, scientists must assign each specimen's map coordinates by hand. But ORNIS, which is not related to the European system of the same name, will assign and update geographical information for each specimen automatically.

"We are using the same concept as the other databases, but making it a lot more efficient," says A. Townsend Peterson, leader of the ORNIS project and curator of ornithology at the Natural History Museum and Biodiversity Research Center of the University of Kansas in Lawrence. "Georeferencing is an enormous task," he explains.

The ORNIS team will also write software to check for errors in the collections. For instance, one program will check for specimens that have been collected suspiciously far from other, similar ones, as such outliers might have been misidentified by their collectors.

Peterson says ORNIS will aid research in other fields by sharing its software. "Other networks can grab the tools we are going to develop and plug them right into their own systems," he says. ■



Birds on the wire: an online resource will collate information on a range of species.



Brain power: Susan Hockfield will become the first female head of MIT in December.

Neurobiologist scores a double first to take MIT presidency

Washington Breaking a 143-year run of male leadership, the Massachusetts Institute of Technology (MIT) in Cambridge has chosen Susan Hockfield to be its sixteenth president. She will take office in December.

A neurobiologist and current provost of Yale University in New Haven, Connecticut, Hockfield will also be the first MIT president with a background in life sciences.

Hockfield graduated from Georgetown University School of Medicine in 1979 with

a degree in anatomy. In 1985, after stints at the University of California, San Francisco, and Cold Spring Harbor Laboratory in New York, she joined Yale's faculty, where she currently studies the development of mammalian brains and the brain cancer glioma. "We will all miss her," Yale president Richard Levin said in a 26 August letter to faculty members and students.

She will replace Charles Vest, a mechanical engineer who announced his retirement last December after 14 years as MIT's president.

Tiny telescope takes long view to discover big planet

London A graduate student scored a victory for the little guys last week by discovering a Jupiter-sized planet some 500 light years from Earth using a telescope just 10 centimetres in diameter — smaller than many amateur astronomers have in their backyards.

The discovery was made by Roi Alonso, a student at the Astrophysical Institute of the Canary Islands on Tenerife, Spain, and part of an international team running the Trans-Atlantic Exoplanet Survey (TrES). The planet, which lies in the constellation Lyra, isn't the biggest or farthest planet yet discovered, nor the first to be detected by the transit method, which looks for a slight dip in a star's brightness as a planet passes across it. But it

is the first pay-off for the TrES survey, which uses three small telescopes in the Canary Islands, Arizona and California to search 12,000 bright stars for planets. Researchers will try to confirm the existence of the planet, dubbed TrES-1, using the 10-metre Keck Telescope in Hawaii.

♦ <http://arxiv.org/abs/astro-ph/0408421>

Small price to pay for drug-company lawsuit

New York Pharmaceutical company GlaxoSmithKline (GSK) is to pay \$2.5 million to the state of New York to settle a legal challenge over its alleged suppression of negative results from clinical trials.

New York's attorney-general, Eliot Spitzer, launched the lawsuit in June, citing an internal GSK memo that outlined plans to "minimize" the commercial impact of results from trials of the antidepressant Paxil (see *Nature* **429**, 589; 2004). Spitzer argued that the trials showed that Paxil might increase the risk of suicide in depressed young people, yet GSK continued to promote the drug to physicians for use in adolescents, he said.

GSK denied the allegations at the time and described them as "unfounded" when announcing the settlement on 26 August. The company says it is making the payment

to avoid the high costs and time required to defend itself against the suit. GSK has already agreed to make all data from clinical trials for marketed drugs available to the public online.

Alleged spy sues Japanese government

Tokyo Takashi Okamoto, an Alzheimer's disease researcher who was recently accused of economic espionage in the United States, is back in the courts for the third time this year — this time of his own accord.

Okamoto is suing the Japanese government for ¥42.9 million (US\$400,000) for mental anguish suffered when the government held him for 57 days earlier this year, pending a Tokyo High Court decision on whether to extradite him to the United States (see *Nature* 430, 960–961; 2004). The court decided that the justice ministry's case was not sufficient to merit extradition.

Okamoto will appreciate the money if another Tokyo court case goes against him. Former friend and cancer researcher Hiroaki Serizawa is suing him for \$770,000 in damages — for legal fees and the loss of his research career — which he says he suffered after Okamoto allegedly involved him in the events that led to the economic espionage case.

China's giant reserve panders to fussy breeder

San Diego China is setting aside a 1,200-square-kilometre tract of land for the endangered giant panda (*Ailuropoda melanoleuca*), officials announced last week. The Baoxing Jiabin Mountain Giant Panda Ecological Tourism Zone will enlarge the Fengtongzhai giant panda nature reserve in southwest China's Sichuan province, and give its 140 giant pandas room to spread out. The scheme is expected to cost 180 million yuan (US\$22 million).

Megan Owen of the Office of Giant Panda Conservation at the San Diego Zoo in California says the plan should enable isolated pockets of pandas to meet up and mate. "It's good to get the gene flow going," she says. Pandas are notoriously coy and difficult to breed in captivity.



Tourist dollars will help pay for the pandas' care. "If it's done responsibly, it's a great way to raise money and support their reserves," says Owen.

Cash pours in to protect Brazil's rainforests

Washington Brazil is to receive a loan of up to US\$1.2 billion from the World Bank over the next four years to help integrate environmental-protection measures into its plans for economic development.

An initial loan of US\$505 million, which was approved on 24 August, together with two other planned loans, will be used to

improve environmental management of critical areas such as tropical rainforests and savannah.

The bank says that the loans make economic sense because Brazil's economy relies heavily on its natural resources. The destruction of these resources by commercial activities, such as cattle ranching, currently costs the country as much as 4% of its annual gross domestic product, say bank officials.



Crunch time for Kyoto

Only Russia can rescue the global agreement on climate change. So why aren't Russian climate scientists speaking up? Quirin Schiermeier and Bryon MacWilliams report from Moscow.



Reach for the seatbelt in a Moscow taxicab, and you are likely to be reprimanded by the driver. “We’re in Russia,” he might say — his way of letting you know he feels insulted by your low regard for his driving skills.

Although the cabbie’s attitude may surprise first-time visitors, it’s old hat to David King, chief scientific adviser to the British government. King is well aware that attitudes in this former superpower are heavily influenced by its dented pride and need for respect.

Yet even he was taken aback at a recent Moscow summit with Russian climate researchers. King and a delegation of British experts had come to the 7 July workshop at the invitation of the Russian Academy of Sciences. But it soon became clear that the agenda had been hijacked by some of Russia’s most vocal critics of the Kyoto Protocol on climate change, the international agreement to reduce global warming, which Russia has yet to ratify. The UK delegates reacted with dismay as several unscheduled speakers got

up to state that man-made global warming was a myth.

King remains diplomatic. “I am very disappointed that this event wasn’t as successful as it should have been,” he told *Nature*. Michael Grubb, a climate-change expert at Imperial College London and a member of the UK delegation, is more blunt. “We walked into a trap,” he says.

That a small group of treaty-objectors could co-opt a high-level scientific meeting illustrates Russia’s unusual situation. Unlike Europe and the United States, where most scientists strongly support efforts to limit greenhouse-gas emissions, Russia is under virtually no pressure from its scientific community to take steps to avert climate change. While the majority remains silent, a small group within the Russian Academy of Sciences speaks with nationalistic fervour about the need to avoid restrictions on the Russian economy. In post-soviet Russia, where economic hardship and growing nationalism are everyday realities, it is difficult for anyone to speak out against them.

The Kyoto agreement, drafted in 1997, requires industrialized countries to cut their greenhouse-gas emissions to 5% below 1990 levels by 2012. It has been ratified or accepted by 124 countries, which collectively account for 44% of the industrialized world’s emissions. But 55% is needed for the treaty to take effect, and, after the United States withdrew its support in 2001, Russia, with its 17.5% share of emissions, became the treaty’s last remaining hope.

Doing a deal

President Vladimir Putin’s statements on the matter have been ambiguous. In May he told European Union leaders that he might be willing to sign in exchange for their support of Russia’s application for membership of the World Trade Organization. But neither Putin nor his ministers have shown any real enthusiasm for the treaty beyond its use as a negotiating tool, and it now seems that a handful of anti-treaty Russian scientists and economists are busy in the background setting the stage for the Kremlin ultimately to reject it.

A. NATRUSKIN/REUTERS



Meltdown? Yuri Izrael (far left, with President Putin) and other powerful voices in the Russian Academy of Sciences (above) oppose Russian ratification of the Kyoto Protocol.

Two powerful advisers to Putin are spearheading this opposition. One is Yuri Izrael, the 74-year-old director of the Moscow-based Institute of Global Climate and Ecology. His age and political leanings have led some of his opponents to call him a “fossil communist fighting for fossil fuel”. The other is Andrei Illarionov, 40, Putin’s top economic adviser and a staunch opponent of any government interference in the economy.

Economic argument

While Izrael says he opposes the Kyoto Protocol mainly on scientific grounds, Illarionov argues that it threatens Russia’s wealth and development. He says that Russia’s rapidly growing economy will soon produce more emissions than the treaty allows, forcing it to buy emission rights on the international market. And as the world’s largest greenhouse polluter, the United States, has backed out despite its much stronger economy, it is even more unreasonable to expect Russia to sign the Kyoto treaty, Illarionov argues.

Observers worry that Putin will find this argument persuasive. Illarionov’s influence is thought to be extensive, and even his critics admit that he is an eloquent and sharp-witted lobbyist with a solid understanding of science. “Not many countries have such an intelligent expert dealing with climate issues at this high level,” says Grubb.

Now the Russian Academy of Sciences has

been drawn into the fray. Izrael, a prominent academy member, was accused of stage-managing last year’s World Climate Change Conference in Moscow to promote his agenda (see *Nature* 423, 792; 2003). Shortly after the conference, Putin asked the academy to reassess the risk of man-made climate change and the effectiveness of the Kyoto Protocol.

Observers considered the move to be politically motivated, as all the key scientific issues had already been addressed in great detail by the Intergovernmental Panel on Climate Change (IPCC), of which Izrael is a vice-president. The last IPCC report in 2001, *Climate Change 2001: The Scientific Basis*, quoted extensive evidence that anthropogenic greenhouse warming is real and could dangerously alter the climate.

But Izrael has challenged this finding. In May, he put out a two-page memorandum based on a series of meetings with selected academy members that claimed “a high level of uncertainty as to whether the rise in temperature [over the past 100 years] was in fact due to human activity”. The Kyoto Protocol has “no scientific basis”, the report concludes.

But Illarionov’s participation in the seminars, and concerns that Izrael’s selection of experts was biased, have led observers to doubt the independence of these conclusions. “I agree there are many uncertainties,” says Igor Mokhov, a climate modeller at the Obukhov Institute of Atmospheric Physics in Moscow. “But no serious analysis can deny the existence of a profound anthropogenic influence on global warming.”

Izrael insists that the academy’s mission is purely scientific and that the treaty itself is politically motivated. “The Kyoto Protocol came about because there is big money being spent on it,” he told *Nature*. “But everyone has forgotten about the climate, and is focusing on how best to trade emission rights and earn money. They’re just deluding themselves.”

The July workshop was to have been an informal exchange of ideas about climate-change research. But unknown to the attendees, Izrael, one of the meeting’s organizers, had invited a group of known climate-change sceptics, including Richard Lindzen, a meteorologist at the Massachusetts Institute of Technology in Cambridge, who is widely regarded as the ‘guru’ of global-warming doubters. Izrael added them to the list of speakers only at the last minute.

This led to complaints by King and other British delegates. “The workshop was an excellent opportunity to discuss the important issue of climate change,” King says. “But there were some very unfortunate last-minute changes made by the organizers.”

In response, Illarionov accused Britain of attempting to force governments against their will to ratify the Kyoto Protocol. “Unfor-

tunately, it is a war. War against the whole world and, in this case, against Russia,” he said at a press conference after the meeting.

The British Office of Science and Technology is playing down the incident, calling it a “storm in a teacup”. But on the quiet, scientists inside and outside Russia worry that it is a sign that the Russian Academy of Sciences is lending its considerable weight to Kyoto-hostile forces in the country.

Cold comfort

This is in marked contrast to the academy’s behaviour in communist times, when concerned members successfully opposed several government proposals they considered misguided, including a plan to re-route large northward-flowing Siberian rivers with a series of controlled nuclear explosions. But in the current climate debate, few within the academy are willing to voice their objections to Izrael’s stance on the treaty.

There is also no public pressure on the government. Climate change is of little concern in a country in the middle of a painful social and economic transition. “Most every Russian family simply has much more immediate problems with adjusting to the new life,” says Mokhov.

And warming doesn’t sound such a bad thing to residents of the coldest country in the world. Indeed, there is a widespread notion, held even by some scientists, that a slightly warmer climate would actually help the country to save energy and produce better harvests.

An upside to climate change is not out of the question, although the reality is likely to be more complicated, says Vladimir Kotlyakov, director of the Russian Academy of Science’s Institute of Geography in Moscow.

Melting permafrost could damage roads and pipelines, for instance. Several institutes of the academy plan to join forces to examine the likely social, economic and physical consequences of a northward shift of climate zones. The results could

help Russian scientists find their voice on the climate issue, says Kotlyakov.

Even this summer’s workshop may have helped by raising awareness among Russian experts about the urgency of the situation, suggests Georgy Golitsyn, director of the Obukhov Institute. “Many of us were quite impressed at the level of preparation Britain is already taking with regards to a changing climate,” he says.

As Russia contemplates what it will do, patience should be the name of the game, says Grubb. “Disputes within the Russian Academy of Sciences are not something that foreign comment will help resolve,” he says. “Every attempt to play hardball with the Russians will backfire.” ■

Quirin Schiermeier is *Nature*’s German correspondent; Bryon MacWilliams is a freelance writer in Moscow.

Brothers in art

Piers Coleman is a theoretical physicist, his brother Jaz a musician with an unusual pedigree. Together, they want to break down boundaries between science and the arts. Sarah Tomlin attends their latest concert.

On almost every street corner in Prague, someone is handing out flyers for a concert in one of the city's historic churches. In one, it's a recital of Mozart; in another, Vivaldi. But on this balmy July evening, I join a crowd of physicists at the Bethlehem Chapel who have come to hear something less familiar: music by a contemporary composer, whose muse tonight is the science of quantum matter.

First, we are treated to a lecture by Tony Leggett of the University of Illinois at Urbana-Champaign, who shared the 2003 physics Nobel for his theories on the behaviour of superfluid helium-3. Then the overhead projector is removed to make way for two harps, a double bass, violin and accordion. A quintet of young Czech musicians takes to the stage. Against a painted backdrop of musical scores from the fifteenth century, melodies from the violin and accordion intertwine to create *Music of the Quantum*.

The concert is the brainchild of the brothers Coleman — one clad tonight in a short-sleeved shirt and tie, the other in a long black jacket over jeans and sandals. The sober-looking Piers is a theoretical physicist at Rutgers University in New Jersey. At 46, he is the elder by two years, but looks more boyish than his bohemian brother. Jeremy 'Jaz' Coleman is composer-in-residence with the Prague Symphony Orchestra. But to many fans, he is better known as leader of the experimental rock group Killing Joke, formed in 1979 from the embers of the British punk scene.

Music of the Quantum was commissioned by the Institute for Complex Adaptive



Renaissance men: Jaz (left) and Piers Coleman at the European premier of *Music of the Quantum*.

Matter, a network of researchers that aims to advance condensed-matter science (see 'Small science thinks big', overleaf). It premiered in 2003 at Columbia University in New York, on the weekend that the United States invaded Iraq. This resulted in a disappointing attendance, as most New Yorkers remained glued to their TV sets.

The brothers are hoping for a more enthusiastic reception in the Czech capital, where the music was composed. Earlier in the day, in a wide-ranging discussion over lunch in Prague's old quarter, they talk about their motivations. "Whether it is science or art, the idea is to make the world a more beautiful place," says Jaz. For Piers, the ultimate goal is to communicate ideas about physics to a wider public. "Scientists have a duty to explain to those who are footing the bill what makes our hearts beat faster," he says.

The da Vinci road

The brothers' respective interests took them down separate paths from an early age. "This is the first time we've done anything together since we were kids," says Piers. They both recall a happy childhood in Cheltenham, UK, nurtured by an English father and a half-Bengali mother — both of them teachers — who encouraged the boys to aspire to Leonardo da Vinci's Renaissance thinking.

For Jaz, school proved a less harmonious environment. He was bullied because of his mixed-race background, but reckons that these hard knocks have served him well. "My father always said it was good training for the

music industry," he says. Jaz gave up on conventional schooling at 14, and three years later formed Killing Joke.

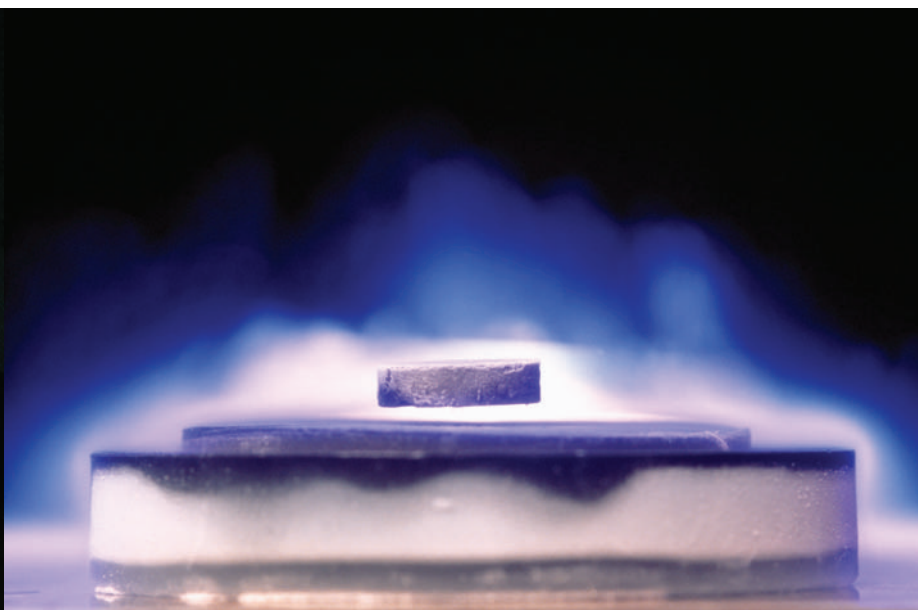
While Piers studied physics at the University of Cambridge and then at Princeton University in New Jersey, Jaz lived the stateless existence of a touring rock musician. Listening to Killing Joke has always been an intense experience — 'doom-laden' is a common description of their heavy, rhythmic sound — and Jaz soon acquired a reputation for eccentric behaviour. In 1982, he suddenly disappeared to Iceland, fearing an imminent apocalypse.

The conflagration never came, and over the subsequent years, Jaz has done his best to live up to his parents' Renaissance ideals. In 2003, Killing Joke released a critically acclaimed eleventh studio album. In parallel, Jaz has built a career as a classical composer. He delights in crossing boundaries, whether composing contemporary music for Maori singers and Czech folk groups or collaborating with Vanessa-Mae on a Disney film soundtrack. He has recorded with artists as diverse as Sarah Brightman and Mick Jagger, and is now working on his second opera.

Piers' career has been more conventional, even though he would love to be able to jump between scientific disciplines in the same way that Jaz leaps from one musical genre to another. "I'd like to think it is possible," Piers says — but concedes that the specialization required of today's scientists makes this an unlikely dream. In tutoring his students, however, Piers tries to encourage a broader outlook, telling them that "equations are like haiku", with their own poetic beauty.



Concerted effort: Tomas Damec (above) and Veronika Vachalova wrestle with physics.



Not antigravity, but just as wonderful — a magnet levitates above a superconductor.

For Piers, the collaboration with his brother represents an unusual opportunity to meld science and art. He first got Jaz interested in physics by sending him *The Dancing Wu Li Masters*, a 1979 book by Gary Zukav that sought to engage the public with theoretical physics by drawing parallels with Eastern philosophy.

The impossible dream

In keeping with this introduction, Jaz has a spiritual approach to science, and is drawn to outlandish ideas. He is fascinated, for instance, by anything to do with antigravity. “As you can see, we have great discussions,” says Piers, laughing. He has told Jaz that antigravity cannot exist, but argues that what science can do — levitating a magnet above a superconductor, for instance, as he shows in a movie during the concert — is just as wonderful. “If you want to do great science, or great art, you need a dream of what is possible,” Piers concludes.

Aside from a chance to collaborate with his brother, Jaz was attracted by the ambitious brief that Piers gave him. Each of the composition’s three movements tries to embody different concepts: emergence and broken symmetry, phase transformation and criticality, and the duality of the quantum world.

Criticality is a tough concept to explain in words, let alone music. Physicists know that new forms of order can develop within matter through phase transformations, which drastically change the material’s properties. Examples include the emergence of magnetism or superconductivity at certain temperatures. Matter that is close to such a transformation is said to be in a critical state, and Piers wanted the movement to convey its dynamic behaviour.

“I am forever trying to understand my brother and his work,” says Jaz. “I’m a musician and composer, so I love numbers.” But when it comes to theoretical physics, Jaz admits that Piers might as well be speaking Chinese. So they communicate in visual terms. For the final movement on quantum duality, Piers wanted a piece from which the listener can either hear two things at once, or by listening carefully just one or the other. So he urged Jaz to create the musical equivalent of an Escher drawing.

Given his own dual musical existence, Jaz may be the ideal composer to write about the quantum world. For him, the rock band is a collective, visceral experience, whereas the life of a classical composer is solitary and cerebral. Killing Joke is about getting “the anger and sickness out of me,” he says. But when it comes to his classical compositions, Jaz is a romantic purist, stressing melody above all else.

Jaz also divides his time between contrasting environments. He spends half the year in Europe, immersed in the cultural intensity of Prague, retreating to a remote island off the coast of New Zealand for the remainder. There he is building a house of his own design, which will be powered by renewable energy.

Raising the bar

In writing *Music of the Quantum*, both Jaz and Piers wanted to engage non-physicists with the beauty of its underlying scientific concepts. “The idea of these concerts is to inspire, to make people think in different ways,” says Jaz.

But the two performances can’t claim to have achieved great outreach. Half of the audience in New York were scientists, the rest

“Scientists have a duty to explain to those who are footing the bill what makes our hearts beat faster.”

— Piers Coleman

Small science thinks big

Condensed-matter physicists have an image problem. Their discipline regularly snares Nobel prizes and dominates lists of highly cited papers, yet the public perception of physics is all particle colliders, black holes and string theory.

In August, the exclusive Colorado ski resort of Aspen hosted a one-day outreach workshop for condensed-matter physicists determined to boost their low public profile. "Who realizes that condensed-matter physics is half of all physics?" asks David Pines, co-director of the Institute for Complex Adaptive Matter (ICAM), formed as an 'institute without walls', to bring together scientists studying the properties of condensed matter, in disciplines from materials science to biology.

ICAM, which organized the Aspen workshop, also aims to grab for condensed-matter physicists a share of the limelight enjoyed by cosmology and particle physics. Pines points to the example of the tiny string-theory community, which has captured the public imagination — in part through the efforts of skilled communicators such as Brian Greene of Columbia University in New York.

A particularly striking example was the list of ten physics problems for the next millennium, released by string theorists in 2000. The list was picked up by the media, and touted as the new physics frontier — even though it did not contain a single question to do with condensed matter.

Following the workshop, ICAM physicists began to draw up their own list of big questions. A preliminary version includes: Is quantum computing feasible? Can statistical mechanics be applied to living cells? Why don't glasses flow?

But some doubt the wisdom of this exercise. "It's relatively rare in condensed-matter physics to predict discoveries. It's a field where you fall over them by accident," says Tony Leggett of the University of Illinois at Urbana-Champaign, who shared the 2003 physics Nobel for his work on superfluidity.

Pines sees this as a virtue, not a problem. The small labs involved in condensed-matter physics aren't organized like the huge teams that work on particle colliders. But that's no reason to ignore the bigger picture, he says.

Others argue that it's unrealistic to expect the public to get as excited about how everyday materials behave as they do about theories that try to explain the Universe. "Condensed-matter physics is an acquired taste," says Neil Mathur of the University of Cambridge, UK.

For many years, this obscurity wasn't a major concern — applications in electronics and other industries kept the cash flowing. But with big US industrial labs being downsized or closed, the field may now have to appeal to the public imagination.

ICAM's final list will be taken to five town meetings being organized by the American Physical Society. These gatherings will provide input for a decadal review of condensed-matter and materials physics being drawn up by the National Research Council. In past reviews, the fundamental excitement of condensed matter was buried in talk of applications. Now it's time to "define a set of questions that are enduring", says Alan Hurd, a condensed-matter physicist at the Los Alamos National Laboratory in New Mexico.

♦ <http://frontiers.physics.rutgers.edu>



Take note: Piers Coleman discusses the physics concepts that inspired the music.

what Jaz calls "science groupies". And owing to restrictions surrounding the use of the venue, the Prague event is not open to the public. Instead, it has been organized as a cultural evening for those attending a meeting on condensed matter, hosted by the European Physical Society.

Piers is disappointed, but hopes that his example will inspire other condensed-matter physicists to reach out to a wider public. "It's a controversial issue," he says — some of his colleagues believe that the beauty of science should emerge naturally from the work, and doesn't need to be packaged for a wider audience. "How élitist," Jaz observes.

In the Bethlehem Chapel, Piers introduces each movement with a brief lecture on some of the physics concepts that inspired

the work, leaving the audience with a final thought or image — a snowflake to represent broken symmetry, for instance.

Jaz is up next. He's a very physical conductor, often moving his whole body, whipping round his shoulder-length hair, with the music. "The three movements are essentially dances," he explains.

Music of the Quantum is uplifting, and the musicians are clearly talented. Piers' talks before each movement are refreshingly wide-ranging and entertaining. Some physicists might baulk at the reincarnation of Schrödinger's cat as the cartoon character Garfield, but Piers says that his 14-year-old son loved it.

Are the physical concepts successfully expressed within the harmony? I'm not

entirely convinced, so I seek the views of scientists with a musical bent in the audience. Michelle Prevot, a chemist from the Max Plank Institute of Colloids and Interfaces in Potsdam, Germany, plays the clarinet and flute. She was impressed by the skill of the musicians and intrigued by the unconventional nature of the quintet — using the contrasting sounds of a violin and accordion to carry overlapping melodies. But Bedrich Velicky, a theorist at Charles University in Prague, and also an amateur musician, felt that the three movements failed to reflect the depth of the concepts involved.

Perhaps physicists are too close to the subject matter — the music was composed for a wider public, after all. And Piers hopes that this goal won't be lost forever. There are plans for a CD and a website, which will include sound and film clips from the concerts, along with interactive quantum experiments and soundbites from interviews conducted with audience members from the New York performance.

The website should be finished by early autumn, but that will depend on the brothers' other commitments. "In parallel with this, serious physics is going on in my life, and serious music is going on in Jaz's," he says.

For his part, Jaz seems happy to be kept busy. "There's no slowing down in our family," he says. It's about "hurtling to the grave, grinning as we go".

Sarah Tomlin edits News Features for *Nature* from New York.

For more news and analysis go to
news @ nature.com
www.nature.com/news

Biosecurity must be internationally supervised

US restrictions on cooperation are hampering legitimate microbiological research.

Sir— In his Commentary article “Isolation is not the answer” (*Nature* 429, 603; 2004), Thomas May warns that US regulations on bioterror-related research in the United States may lead to scientific isolationism.

We can confirm that national regulations do affect scientists in other countries. A survey we conducted recently in Germany revealed that US regulations are having an adverse impact on German microbiological research.

In spring 2004, we sent questionnaires to all German academic medical and veterinary institutions that are, according to their websites, working in the area of microbiology. From a total of 67 respondents (response rate 33%), 47 commented on changes in cooperation with US partners. Of these, 21 stated that access to US culture collections or

microorganisms had become “worse” (11 respondents) or “much worse” (10 respondents) since 11 September 2001. This unfavourable development has had a direct negative effect on microbiological research in Germany: 13 respondents stated that there had been measurable impacts, such as delays in some research projects, the need to switch to different organisms or technologies, or the need to cooperate with other partners. Two microbiologists reported that projects had to be stopped, or could not begin, because of reduced cooperation with US partners.

A certain response bias cannot be ruled out. Researchers affected by US regulations are more likely to have responded to the questionnaire than those not working with pathogens or not cooperating with international partners.

The responses nevertheless indicate that failure to harmonize international biosecurity and biosafety measures may well hamper legitimate scientific research outside US borders.

Enhanced supervision of microbiology and related research is much needed, but it should follow internationally agreed rules (see J. D. Steinbruner and E. D. Harris *Issues Sci. Technol.* 47–54; Spring 2003). As the US National Research Council and the UK Royal Society have suggested, harmonized international supervision would be better suited to accommodate both scientific and security concerns.

Jan van Aken, Stefan Johannsen, Regine Kollek

University of Hamburg, Falkenried 94, D-20251 Hamburg, Germany

Linnean Society backs Godfray on use of web

Sir— Your Editorial “Ignorance is not bliss” (*Nature* 430, 385; 2004) notes that Charles Godfray “argues that taxonomy must emerge from museums to become a web-based information science”. It continues: “Some initiatives of this ilk are under way, but the call has been short-sightedly rejected by much of the taxonomic community, notably the Linnean Society of London.”

I was president of the Linnean Society from 2000 to 2003. During this period the society submitted written evidence to the Inquiry into Systematic Biology and Biodiversity held by the House of Lords Select Committee on Science and Technology. This evidence was published in *What on Earth? The Threat to the Science Underpinning Conservation: Evidence* (HL paper 118 (ii); 2002).

On pages 124–125, the following statement occurs as part of the Linnean Society’s evidence (all of which was formally approved by its council): “Professor Charles Godfray FRS ... argues powerfully and persuasively for a major sea-change in taxonomy whereby the systematics of all groups of organisms would become a single web-based resource ... His proposal would have the particular advantage that at last, taxonomic information would become easily available ... This will be essential if real and effective progress is to be made in the conservation of biodiversity in the UK.”

The Linnean Society therefore does not

reject but supports the initiatives that have been proposed by Charles Godfray.

David Smith

13 Abbotsford Park, Edinburgh EH10 5DZ, UK

Need for economists to set global priorities

Sir— In May, the Copenhagen Consensus (www.copenhagenconsensus.com) brought 38 of the world’s top economists to Denmark to make a prioritized list of solutions to global challenges. In his critical judgement of the consensus, “Seeking a global solution” (*Nature* 430, 725–726; 2004), Jeffrey Sachs misunderstood the project in three fundamental ways.

First, Sachs asserted that, because the world has promised to spend much more than the US\$50 billion discussed by the Copenhagen Consensus, prioritization is unnecessary. But more or less money would not alter the project’s outcome. If more than \$50 billion were marshalled, the consensus list would simply show where the extra money should be directed.

The claim that the developed world has promised to substantially increase development assistance should be treated with caution: since 1970 the United Nations has wanted such spending to double as a percentage of GNP, but it has fallen substantially since then (see www.worldhunger.org/articles/global/Intconforaid/develfinance.htm). It seems unrealistic to assume that the resources allocated will be so large that we will not

have to prioritize. The \$50 billion was chosen as an example of a realistic commitment to additional spending.

Second, Sachs’ criticism that the Copenhagen Consensus consisted solely of economists missed the very point of the project. Economists have expertise in economic prioritization. It is they and not climatologists or malaria experts who can prioritize between battling global warming or communicable disease. Of course, all economic estimates are based squarely on the best natural-science models.

Third, in discussing why efforts to handle climate change ended up at the bottom of the Copenhagen Consensus list, Sachs left the distorted impression that all climate economists believed the relatively high carbon-tax proposal should have been given a higher priority. In fact, two of the three climate economists explicitly found the costs exceeded the benefits. The expert panel endorsed this view.

The Copenhagen Consensus was the first project to prioritize the major challenges facing the world. Morally we must focus on the best priorities first, or we do less good for humanity. Prioritization means not everything is done first. This often makes good people fret because ‘we should do everything’, although we do not and cannot.

That a prominent economist such as Sachs should posit such flawed arguments underlines the case both for the Copenhagen Consensus and for a continued public discourse on prioritization.

Bjørn Lomborg

Copenhagen Consensus, Linnegade 18, DK-1361 Copenhagen K, Denmark

Politics, morals and embryos

Can bioethics in the United States rise above politics?

**George J. Annas and
Sherman Elias**

Bioethics in the United States reflects US culture and tends to be pragmatic, market-oriented and insular. Add embryo politics to this mix and, over the past few years, the result has been a bioethics that has become so narrow and self-absorbed as to be virtually irrelevant to the rest of the world. Not all the blame for this can be placed on President George W. Bush's political agenda for his President's Council on Bioethics, now in its third year of operation, but much can. The council has made public bioethics the servant of politics by pursuing a narrow, embryo-centric agenda. More remarkably, although the attacks of 11 September 2001 changed almost everything in the US government, the bioethics council — and bioethics in general — were strangely unaffected.

Both the president and the chair of the council, Leon Kass, have cited Aldous Huxley's *Brave New World* as justification for the council's work. The threats portrayed in this novel — artificial reproduction of humans in state 'hatcheries' and drug-induced contentment — are real, but preventing them is not the world's only, or most important, bioethics problem. Both domestic and global access to healthcare, the commercialization of science and medicine, pharmaceutical pricing, conflicts of interest, gene patenting and international research rules merit at least as much attention.

It is understandable, but not acceptable, that a neoconservative bioethics council would have nothing to say about access to healthcare by the tens of millions of uninsured and underinsured Americans. But after 11 September, the president and Congress immediately set about writing new laws that had profound ethical implications for medical research, including the USA Patriot Act, the Bioterror Act and the Bioshield Act. It is almost beyond comprehension that the council had nothing to say about any terrorism-related medical research issues: not about classified biotechnology research; not about the attempted smallpox vaccination of 500,000 US health workers and children; not even about the testing of bioterrorism countermeasures on humans.



Ethical questions about frozen embryos dominate US politics.

Of course, an administration that has declared an open-ended war on terror might not want the bioethics council involved in these discussions. The Bush administration certainly does not want to suggest to US citizens that George Orwell's *1984*, with its perpetual war, doublespeak and government reliance on fear, is as relevant to contemporary bioethics as *Brave New World*. In the war on terror, the nonpartisan US National Research Council has reasonably concluded that controlling biotechnology for terrorism and warfare will require international cooperation¹. So will meaningful bioethics.

We believe it is the narrow focus of US bioethics, both geographically and philosophically, that permitted terrorism and war to be placed ethically 'off-limits'. The international language of ethics is the language of human rights, and international human rights law in particular, but US bioethics retains an isolationist worldview. From the bioethics council's perspective, using the global language of human rights would have

been counterproductive, because embryos do not have the same status as human persons in any human-rights document. But whether it is the failure of US physicians at Abu Ghraib to stop the abuse of Iraqi prisoners there, or the possible ethical shortcuts taken by Korean cloners in obtaining human eggs, a global bioethics based on human rights would demand respect for the human dignity and rights of people — not just embryos. Bioethics in the United States has been noticeably silent on these issues².

Had the United States adopted an international human-rights perspective on bioethics, it too could have joined the UN proposal of August 2001 by France and Germany to draft a treaty to outlaw human reproductive cloning, thereby protecting children. US joint sponsorship of this proposal could have led to the world's first bioethics treaty. Instead, embryo-centric bioethics led the United States to effectively kill the proposal by insisting that any cloning treaty simultaneously outlaw the creation of embryos for research purposes.

A full understanding of the President's Council on Bioethics requires recalling its original political agenda. The council was announced by President Bush during a speech on 9 August 2001 in which he said he

would limit federal funding for embryonic stem-cell research to stem-cell lines that had already been created "where the life-and-death decision has already been made", and would provide no "taxpayer funding that would sanction or encourage further destruction of human embryos that have at least the potential for life". The council's mandate was to "monitor stem-cell research, to recommend appropriate guidelines and regulations, and to consider all of the medical and ethical ramifications of biomedical innovation".

The bioethics council has so far issued four reports and one collection of readings. Only two of the reports contain policy recommendations. The best known is *Human Cloning and Human Dignity* (2002), which called on Congress and the president to ban human reproductive cloning and to have a moratorium on research into cloning for therapeutic purposes. The two most recent reports of the commission focus directly on stem cells. *Monitoring Stem Cell*

Research (2003) is a heroic attempt to provide an ethical justification for the stem-cell policy articulated in Bush's speech of 9 August 2001. The 'ethics' of the Bush position is based entirely on the argument that taxpayer money should not be spent on destroying any more human embryos in the United States — the private sector and other nations can do what they want. This means that public funding is crucial to the argument.

The council boldly attempts to justify this position, saying: "The decision to fund an activity is...a declaration of official national support and endorsement, a positive assertion that the activity in question is deemed by the nation as a whole, through its government, to be good and worthy. When something is done with public funding, it is done, so to speak, in the name of the country, with its blessing and encouragement." This is powerful rhetoric, but wrongly equates politics and morals. It is the criminal law that reflects the nation's minimum morality, not funding decisions³. Congress funds, or refuses to fund, thousands of projects that are of interest to only tiny minorities — sometimes called 'special-interest' groups. Federal funding of these projects, whether tobacco-farming subsidies or new nuclear weapon designs, does not imply that they are "deemed by the nation as a whole...to be good and worthy".

State funding decisions are also political. This November, voters in California will decide whether to amend their state constitution to make embryonic stem-cell research for therapeutic purposes a constitutional right, and to provide US\$3 billion in public funds for stem-cell research. This proposal is excessive, but it is an understandable referendum on the Bush administration's stem-cell policy. It also demonstrates the muddled ethics of the bioethics council, whose overall position would cause it to reject the California initiative, but whose "will of the people" rationale lends support to the idea of voting to determine the morality of this research.

Born with an embryo-centric, anti-abortion and anti-regulation political agenda, Bush's President's Council on Bioethics has repeatedly failed to transcend it. In its latest report, *Reproduction and Responsibility* (2004), the council attempts to come to grips with the ethics of assisted-reproduction technology, but ultimately reverts to embryo protection. A thoughtful national report on public oversight for the assisted-reproduction industry is long overdue in the United States. Unfortunately, this report — unlike much earlier UK and Canadian reports — avoids almost all the tough ethical-policy issues in assisted reproduction by concluding that there is insufficient factual information available to make regulatory recommendations.

"We believe it is the narrow focus of American bioethics, both geographically and philosophically, that permitted terrorism and war to be placed ethically 'off-limits'."



Leon Kass (above) and George Bush have "made public bioethics the servant of politics".



In *Reproduction and Responsibility* (2004), the council also makes the only legislative recommendations it has made since its cloning report, the most important of which are to "prohibit attempts to conceive a child by any means other than the union of egg and sperm; prohibit the use of human embryos in research beyond a designated stage in their development (between 10 and 14 days after fertilization); and prohibit the buying and selling of human embryos". These prohibitions are reasonable, but a more ethically and politically constructive approach would be for Congress to provide both federal funding and meaningful federal regulatory oversight for embryonic stem-cell research.

During the Clinton administration, we supported the US National Institute of Health's position that federal funding should be available for embryonic stem-cell research conducted with 'surplus' *in vitro*

fertilization embryos donated by couples. However, we noted at the time that the rationale for this limitation on the federal funding of embryo research was political — not ethical⁴ (the moral status of a human embryo no more depends on how or why it was created than does the moral status of a child). It is easier for a politician to support research with 'surplus' embryos than with embryos created for that purpose. But if one believes that embryos should never be created or destroyed for research, such activities should be made a crime, not turned over to an unregulated private sector.

The 'surplus embryo' compromise may still be politically possible, having gained the support of 58 US senators (including some who oppose abortion) and former first lady Nancy Reagan, but not until after the November presidential elections. Given the inability of Bush's bioethics council to provide a credible ethical rationale for his August 2001 position, Bush may pay a political price in the November elections, especially from patient advocates and their families.

Bioethics is important in US politics, just as morality is important in law-making, but when bioethics is used primarily to serve an ideological, domestic political agenda, rather than helping to develop a global ethic, it is of little use to anyone but narrow-interest groups. It is too late to reform Bush's bioethics council. Even mainstream bioethicists who accept the reality that federal panels can never be totally divorced from politics find the politics of this council extreme and isolated⁵. In our view, future bioethics panels in the United States must be independent, not 'presidential'. For the benefit of medicine, science and society, it would be better to establish a permanent National Institute of Bioethics as part of the National Academy of Sciences. Its mandate must be broad, and it should adopt international ethics, especially those embraced in the Universal Declaration of Human Rights, so that it can proceed from a global rather than exclusively US perspective⁶.

George J. Annas is at the Department of Health Law, Bioethics & Human Rights, Boston University School of Public Health, 715 Albany Street, Boston, Massachusetts 02118, USA.

Sherman Elias is at the Department of Obstetrics and Gynecology, Feinberg School of Medicine, Northwestern University, 333 East Superior Street, Chicago, Illinois 60611, USA.

1. *Biotechnology Research in an Age of Terrorism* (National Academies Press, Washington DC, 2004).
2. Annas, G. J. *American Bioethics: Crossing Human Rights and Health Law Boundaries* (Oxford Univ. Press, New York, in the press).
3. Fuller, L. *The Morality of Law* (Yale Univ. Press, New Haven, 1964).
4. Annas, G. J., Caplan, A. & Elias, S. *New Engl. J. Med.* **334**, 1329–1332 (1996).
5. Charo, R. A. *J. Law, Med. & Ethics* **32**, 307–314 (2004).
6. *The Medical Profession and Human Rights for a Changing Agenda Handbook* British Medical Association (Zed Books, London, 2001).

When giants walked the Earth

A pedigree of Darwin's well bred English bulldogs.

A Reason for Everything: Natural Selection and the British Imagination

by Marek Kohn

Faber & Faber: 2004. 400 pp. £20

Steve Jones

What's this cult of personality in evolutionary biology all about? There's the Great Leader, Chairman Charles, of course, and various lesser but substantial figures who are also worthy of the occasional parade. But why do we need so many? Experts on chloroplasts or chlorine manage, as far as I know, with living facts, and are not forced to attach them to dead heroes. But there's something in evolution that calls for immortals to whom we plebs must defer.

A Reason for Everything explores the lives of six members of the central committee of the English Evolutionary Party and their hangers-on. With a single exception, the players are toffs to a man, products of famous public schools followed by one of the older provincial universities. Charles Darwin (Shrewsbury and Cambridge) blotted his copybook by spending a student year in Scotland, but of the six discussed here, R. A. Fisher (Harrow and Cambridge), J. B. S. Haldane (Eton and Oxford), John Maynard Smith (Eton and Cambridge), Bill Hamilton (Tonbridge and Cambridge) and Richard Dawkins (Oundle and Oxford), all had a grand English education. Kohn's one great anomaly is Alfred Russel Wallace, who enjoyed a short period at Hertford Grammar School and thereafter had to make do with the University of Life.

Each is given a sympathetic hearing, although one senses that Kohn's patience is tried by the miasma of self-congratulation that surrounded some of the actors in his drama. Fisher claimed that his fundamental theory of natural selection occupied the supreme position among the biological sciences, although others dismissed it as a verbal trick. Fisher's colleague E. B. Ford, a fellow of All Souls College, Oxford, a silly man in a silly place, did some desultory research and spent his later years bemoaning the presence of women in lectures and waiting for ecological genetics to supplant all that dull molecular stuff.

The most remarkable character is the earliest of Darwin's bulldogs (although he lived long enough to worry about air warfare). Wallace's expeditions were followed by a lifetime of devotion to the Great Leader and to a series of ingenious observations, including an attempt to prove, for a bet, that the world was not flat. His elegant demonstra-



A statistical geneticist and the Rajah of Bomb: R. A. Fisher (left) and J. B. S. Haldane.

tion led to death threats (and contempt from his better-off colleagues) but was good science. He turned, alas, to spiritualism and, as so often when scientists use their knowledge of nature to interpret the world of man, abandoned common sense.

Some of his successors were also happy to use Arts Faculty science — sweeping generalizations without the need for facts — when discussing human affairs. Fisher was, like his near-contemporary J. R. R. Tolkien, an undergraduate fan of the Nordic myths. His *Genetical Theory of Natural Selection* became evolution's equivalent of *The Lord of the Rings*: full of gnomish and portentous truths with rather a nasty social agenda lurking beneath (Fisher felt it his biological duty to beget eight children). As Kohn points out, Fisher's followers, like those of Wagner — composer of a musical on the same theme — are obsessed with the fine detail of what the great begetter meant and are still far from sure.

Set against the bearded bigot, the Gandalf-like figure of Haldane is revealed in a rather better light. A daring and often reckless experimenter, he was known in the trenches as the Rajah of Bomb and was pursued by the whiff of cordite throughout his career. He stuck with the Communist party long after his colleagues had abandoned it, and Kohn provides a telling account of Haldane's readiness to support Comrade Lysenko even in the face of powerful evidence against his theories.

Haldane's representative on Earth was, for nearly forty years, John Maynard Smith (who had himself hung the hammer and sickle from his Eton window). Maynard Smith inspired many young evolutionists, and I have seen him hold an entire pub entranced as he discoursed to a group of undergraduates. His interview for this book reveals many details of his life and way of thinking, and shows how conflict has overtaken cooperation as the key to understanding animal behaviour.

In a recent magazine poll, Richard Dawkins, with his trademark hobbit smile, was voted Britain's top intellectual (a welcome kick in the teeth for the new generation of Creationists in our privately funded schools). He, too, spoke to Kohn, but perhaps not for long ("Dawkins, the most public but most private of scientists..."). Darwin's latest bulldog is, it seems clear, not best friends with the Bush regime but, unlike too many others, Dawkins has been scrupulous in not allowing his science to control his politics.

The most ambiguous character to emerge from these pages is Hamilton. He found it hard to make friends and nursed long, Gollum-like resentments in his search for the ring of truth. Kohn suggests that he may have suffered from autism. Autistic or not, Bill Hamilton was an outstanding biologist. Darwin's ability to generalize came from his huge knowledge of plants and animals. Few of his intellectual descendants can tell

a hawk from a handsaw, let alone from an eagle. Hamilton could, and his daring tropical trips gave him the raw material for many leaps of the scientific imagination. His last, published posthumously, argues that the plants with the brightest autumn colours are telling parasites that they are fit and healthy, and to go elsewhere. Hamilton, sad to say, was also a martyr to political vapourings and lobbied for a cracked eugenical Utopia with Margaret Thatcher as Life President and caesarian births banned.

A Reason for Everything is a well-written and carefully researched account of some of the main British players in the world of evolution. Every evolutionist should read it — as a warning against personality cults, if nothing else. Kohn makes it clear that giants walked the Earth in those days. Those days are gone, but after perusing his chapter on the Oxford school of evolutionary biology in the 1950s and 1960s — some geniuses, no doubt, but also a fair sprinkling of prima donnas and right-wing zealots — one can only mutter, through gently clenched teeth, “Thank God!”

Steve Jones is in the Galton Laboratory, University College London, Gower Street, London WC1E 6BT, UK.

Well bred rodents

Making Mice: Standardizing Animals for American Biomedical Research, 1900–1955

by Karen Rader

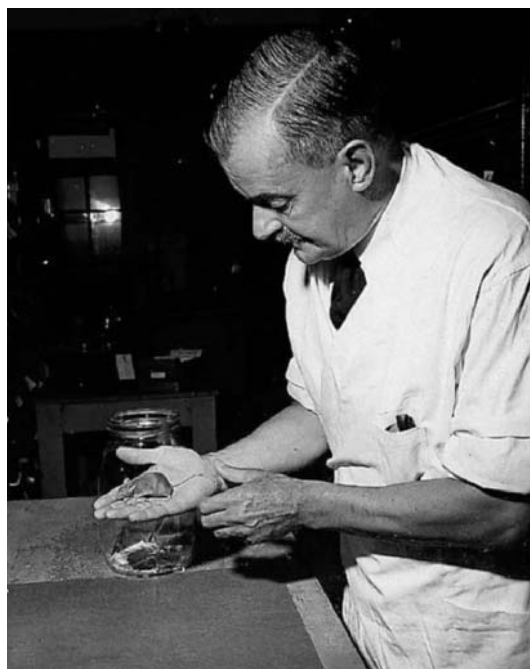
Princeton University Press: 2004. 312 pp.

\$45, £29.95

Michael Festing

Clarence Cook Little started his undergraduate work in genetics at Harvard's Bussey Institute in the class of 1910. This was soon after the field of genetics was established following the rediscovery of Gregor Mendel's classical paper on unit inheritance in peas. Initially Little studied the inheritance of coat colour in mice. In order to produce a pure line purged of recessive genes, and to study the effects of inbreeding, he mated brother and sister mice for several generations to produce DBA, the first inbred strain. He soon realized the importance of using pure-line mice in research, something that many research workers still do not appreciate today. Little's mice have made major contributions to our understanding of cancer and many other branches of mammalian biology. Without them we would probably not have monoclonal antibodies or a technique for knocking out individual genes in order to study their function.

In 1915, Little came to prominence in cancer research when he started a 30-year



A helping hand: the pure-bred mice developed by Clarence Cook Little are a valuable tool for biomedical researchers.

scientific feud with Maud Slye, who claimed, based on studies with outbred mice, that cancer was inherited as a simple mendelian recessive character. Little was right: the cause of cancer is much more complicated, depending on many genes and environmental factors.

Little continued his research at the Cold Spring Harbor Laboratory, but in 1922, aged just 34, he was appointed president of the University of Maine. He subsequently met many of the rich summer visitors to Bar Harbor, including Roscoe B. Jackson, an engineer and joint founder of the Hudson Motor Car Company, who supported some of Little's research projects. With these connections and his strong research interests, Little obtained an even more prestigious appointment in 1925, as president of the University of Michigan.

Four years later, Little set up his own research institute in Bar Harbor on land donated by George Dorr. It was named, with Dorr's approval, the Roscoe B. Jackson Memorial Laboratory, after Jackson, who had died unexpectedly from influenza. Little could not have chosen a worse time. The stock-market crash of 1929 occurred within days and sources of support among the seasonal residents of Bar Harbor dried up. Little was forced to sell mice to keep the lab going, despite a tradition of supplying genetically interesting stock to colleagues free of charge. In subsequent years, many scientists and bureaucrats have found it difficult to understand how an organization could be both a research institute and a supplier of animals. Little emphasized repeatedly that the prime purpose of the laboratory was research, the supply of animals being a

service to the research community.

In 1947 the Jackson Laboratory was destroyed in a fire that killed 14 people and thousands of mice. Little received many offers of help, and research scientists from throughout the United States sent back breeding pairs of mice to re-establish the colonies. By 1949 it had been rebuilt and continued in its role as a centre for mammalian genetics and biomedical research.

Making Mice is not exactly a history of inbred mice. There is no scientific account of the characteristics of such strains, nor of how they have made such a substantial contribution to biomedical research. Nor is it a biography of Little, who did so much to promote them; we learn only indirectly about the man and his background. I suspect that he was driven by his vision of pure-line mice as an essential research tool, and was intolerant of anyone who disagreed with him. Nor is it really

a history of the Jackson Laboratory; there is, for example, a long section on the 'mega-mouse' experiments at Oak Ridge National Laboratory in Tennessee.

Rather, the book puts its three themes — mice, Little and the Jackson Laboratory — into historical context and weaves them together, the chapters following a rough chronology. It shows the tortuous path that finally led to the establishment of this world-renowned institute, and should be of interest both to historians and to scientists who work with genetically defined mice. There are a few misprints, however, and the book is incorrect in implying that rats, mice and birds are now protected under the US Animal Welfare Act.

The Jackson Laboratory has become an internationally recognized centre of excellence for research, teaching, informatics and the supply of genetically defined mice. The case for using inbred mice in research has largely been accepted, except in a few reactionary disciplines, such as toxicology. However, I consider it a major scientific scandal that in 2004, nearly a century after Little first advocated the widespread use of inbred, pure-line mice, more than 80% of scientific papers using the rat still involve genetically undefined Wistar or Sprague-Dawley rats. According to Little: “Just as the purity of the chemical assures the pharmacist of the proper filling of the doctor's prescription, so the purity of the mouse stock can assure a research scientist of a true and sure experiment.” How many rats are being wasted as a result of the inability of many scientists to understand this simple point?

Michael Festing was formerly at the MRC Toxicology Unit, Leicester LE1 9HN, UK.

Into the woods

Encyclopedia of Forest Sciences

edited by Jeffery Burley, Julian Evans & John A. Youngquist
Academic Press; 2004. 2,400 pp. £569, \$875

Colin Tudge

The past few hundred years of industrialization have left the world with a series of what can reasonably be called crises: of climate, food, and cities that are ever more out of control. All of these involve trees, in one way or another. Get forestry right, and a lot more will follow.

There is much to encourage in the *Encyclopedia of Forest Sciences*, which provides an excellent summary of current ideas and practice. The modern science of forestry is often breathtaking. On the grandest scale, satellites can measure the height of forests to within about 30 cm, monitoring the growth and the density of leaves over entire continents — a task that would take lifetimes of work from the ground. The exchange of gases and volatiles from the canopy and roots is measured around the clock and the calendar. Hectares of tropical forest have been starved of most of their water, by covering the ground and draining the rain away, to measure the effects of drought.

From such measures (and a great deal of computer power), the effects of global warming can be modelled. As carbon dioxide increases, so too should photosynthesis (which uses it up), and there is now experimental evidence for this. But temperature should also rise, increasing respiration, which slows growth, and harming trees that are not adapted to a hotter climate. As they die, there is a short-term net release of carbon dioxide.

Global warming can increase rainfall in high latitudes and the tropics yet reduce it in the subtropics. But the effects will be patchy



A green solution: maintaining forests, like this one in Hawaii, can help to counteract climate change.

and erratic; in particular, El Niño is a huge and unpredictable player. Where there is drought, trees will stop growing and again become net releasers of carbon dioxide. More heat means more drying, which means more fires — another major source of carbon dioxide (with more released from the bare ground left behind). If leaf litter is left *in situ*, the chance of fire increases; if it is removed, nitrogen is lost too and soil fertility is reduced. It is a difficult balance to strike, but there is no doubt that trees ameliorate climate change. The world needs more trees.

On a smaller scale, biologists can now sample the DNA in the cambial cells of wild trees, just beneath the bark, so they can track the loss of genetic diversity under different logging regimes. Forestry practice is still, in large part, craft. Sustainable harvesting, the great desideratum, is largely governed by good sense and rules of thumb. This is as it should be; as the world's agriculture shows, we abandon common sense at our peril. But the more science that is deployed in the service of craft and sense, the better.

Forestry has an ambivalent relationship with farming. Both are vital. Both determine landscape. Often they have been in conflict. But they can operate beautifully in harmony through the much neglected but highly promising pursuit of agroforestry. Trees are grown mainly for timber and shade, but they are also — or could be — huge producers of food and fodder in their own right. Agroforestry would be a much more effective employer in the developing world than the spreading estates of soya, coffee and other industrialized commodity crops that are now the rage. And viable agrarian economies have never been needed more, to relieve the swelling cities.

In one crucial respect, foresters now have a significant edge over farmers. Consumers can eat only so much food, and farmers always face the spectre of overproduction, which is only slightly less destructive than crop failure. I view the continuing emphasis on productivity in agricultural science (epitomized by the zeal for biotech) as a kind of lunacy, although it is highly profitable for its controllers.

But there need be no ceiling on the raising of trees. Flourishing forest is a carbon sink. Trees used as fuel are carbon neutral. Much more timber could be used in construction, as protected teak doesn't sag like plasticine when hot, as steel does. Of course, it takes extra energy (and hence more carbon) to turn a tree into usable timber, but it takes 12 times as much to produce steel to do the same job. If carbon taxes are applied to building materials, which surely is desirable, then the market for timber would be insatiable.

Forestry has always been at the heart of human economies and now it seems more important than ever. But bringing its ideas together into a satisfying whole is not easy. This encyclopedia does the job triumphantly. If you are at all involved with trees, you should have a copy. And who isn't? ■

Colin Tudge is a writer based in Oxfordshire, UK. His latest book, on farming, is *So Shall We Reap*.

New in paperback

Galileo's Finger: The Ten Great Ideas of Science

by Peter Atkins

Oxford University Press, £8.99

"This is a charming and ambitious book that I would not hesitate to recommend as a gift for a young person on the threshold of a scientific career." Frank Wilczek, *Nature* **422**, 377–378 (2003).

Eurekas and Euphorias: The Oxford Book of Scientific Anecdotes

by Walter Gratzer

Oxford University Press, £8.99, \$15.95

A book of scientific anecdotes that "cover the entire range of human genius, folly, character and accident". Oliver Sacks, *Nature* **419**, 786 (2002).

The Extravagant Universe: Exploding Stars, Dark Energy and the Accelerating Cosmos

by Robert P. Kirshner

Princeton University Press, £12.95

The story of the acceleration of the Universe "is related with verve and good humour". Sean Carroll, *Nature* **419**, 784–785 (2002).

The Solar Economy: Renewable Energy for a Sustainable Global Future

by Hermann Scheer

Earthscan, £17.99, \$24.95

Safe Food: Bacteria, Biotechnology, and Bioterrorism

by Marion Nestle

University of California Press, £10.95, \$15.95

Midsummer madness

How an unexpected allegory led to the birth of a new subject.

Gautam R. Desiraju

I have always regarded myself as unconventional, taking roads not taken by others. Therefore, in 1978, despite a green card and being armed with a PhD from the University of Illinois, I came home permanently to India. Everyone else I knew was travelling in the opposite direction.

Of course, my idealism had wings of wax, which began to melt in the enervating heat and dust of the fledgling University of Hyderabad. The infrastructure in our labs was abysmal and, literally, state-of-the-ark. There was no sign whatsoever of a single crystal X-ray diffractometer, the main equipment that I needed for my work on crystal structure determination. We operated out of what were little more than sheds. In addition, the scientific system was charmingly rural, and I did not want to understand it. The organic solid-state chemistry I had learnt in the United States seemed better related to life on another planet. I knew I had to visit a laboratory abroad before I began to abandon hope.

The summer of 1982 found me in Cambridge where John Meurig Thomas had just begun a vigorous programme on organic topotactic reactions, in which solid reactants transform to products without loss of crystallinity. I was not prepared for what I would encounter. The Thomas group was large and purposeful, but tackled all their problems in a seemingly relaxed manner. Then there was Cambridge itself, with its restful colleges serenely silhouetted against the mild magic of an English summer. The long days and short nights, with the sky refusing to darken into the familiar ink-black of the tropics, only heightened my exhilaration and anxiety. I worked on topotactic reactions with a sinking feeling. With no equipment back at home, I knew I would soon be reduced to a guest worker in Cambridge. Still, I returned to Cambridge the following summer, but this time my predilection for the unconventional lured me into the company of Simon Kearsley and Noel Thomas — the maverick students of the group — who were working on problems largely unconnected with topotaxy.

Simon and Noel were looking at crystal structures just to see if they could understand

why they were packed the way that they were. But back then, this was not really a topic fit for research. Determining the structure of a crystal was difficult enough; few people worried about how different crystal structures interrelated — it was simply too difficult, and not enough was known. Aleksander I. Kitaigorodskii had said that close-packing of molecules was all that mattered, with this close-packing resulting in low crystal symmetry. Gerhard Schmidt, on the other hand, had tried to ‘engineer’ crystal structures by molecular manipulation. His so-called

Schmidt’s chloro rule. We noticed something rather unusual. Only three of the compounds obeyed the rule, but these were the three that crystallized with higher symmetries, violating Kitaigorodskii’s rule. The three other isomers did exactly the opposite. They did not have 4 Å axes, but they adopted low-symmetry space groups. So, either Schmidt or Kitaigorodskii was correct, but not both, in any particular instance. This meant that violations of the close-packing principle could be rationalized on the basis of effects of specific groups and directional interactions. My excitement

quicken when I realized that by studying these interactions, it might be possible to discern rules for designing organic solids with specific physical and chemical properties — the goal of ‘crystal engineering’. With 50,000 crystals in the CSD (and no prospect of a diffractometer back home!), surely crystal-gazing was going to be a sporting proposition. Of course, this all came much later, and we were happy enough to show our results to John Meurig Thomas who generously suggested that Noel and I publish this work independently.

Our observations that day radically altered the way I looked at science and the directions my research took thereafter. Since then, the work I have done has had a major impact on the growth and development of crystal engineering, but much of it goes back to that single result and its implications. Today, when crystal engineering is described as “one of the principal challenges of modern chemistry” (*Chem. Eur. J.* **10**, 3783–3791, 2004), I recall with amusement the irreverence with which Simon, Noel and I would discuss crystal packing. It was probably our midsummer madness that made us question existing concepts and paradigms. Without quite realizing it, we were supervising the birth of a new subject. ■

Gautam R. Desiraju is in the School of Chemistry, University of Hyderabad, Hyderabad 500 046, India.

FURTHER READING

Desiraju, G. R. *Crystal Engineering* (Elsevier, Amsterdam, 1989)

Desiraju, G. R. *Angew. Chem. Int. Ed. Engl.* **34**, 2311–2327 (1995).

Desiraju, G. R. & Steiner, T. *The Weak Hydrogen Bond in Structural Chemistry and Biology* (Oxford Univ. Press, Oxford, 1999).



Crystal-gazing: a placement at Cambridge had an unexpected influence.

‘chloro rule’ stated that planar aromatic molecules with two chloro-group substitutions tend to form crystals with a short axis of 4.0 Å. This was all that was known about the packing systematics of organic compounds.

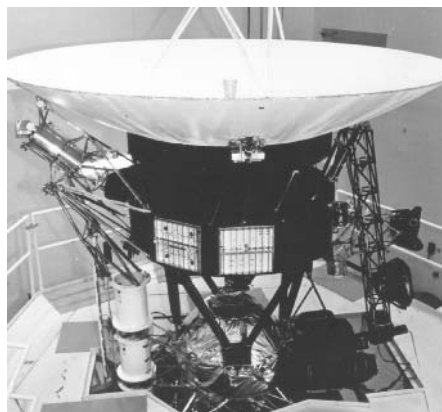
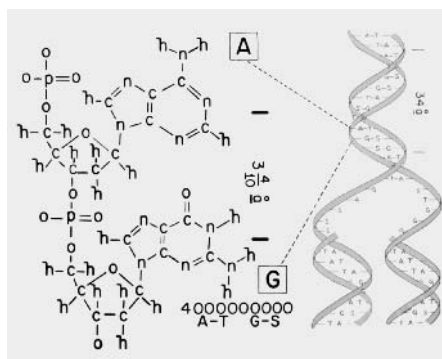
Simon, Noel and I were not even housed with the rest of the group. We occupied a cavernous undergraduate laboratory that lay deserted for those summer months, save for hundreds of models — papier mâché lions, huge wooden cannonballs and metal wire storks sent in by ‘A’-level students from across the world for their practical examinations. The room was piled high with these curiosities and they stared eerily at us during the long evening hours, as we in turn stared at crystal structures. Fortuitously, the Cambridge Structural Database (CSD) was being developed by Olga Kennard and her associates in the very same building at the time, so Simon and Noel were able to retrieve crystal structures using this brand new computational tool.

The turning point in my research came when Noel showed me the crystal structures of the six isomeric dichlorophenols that he had extracted from the CSD to verify

Woodruff T. Sullivan III

Are we alone? Although the Search for Extraterrestrial Intelligence (SETI) has yet to detect a signal, the efforts continue because so little of the possible parameter space has been searched so far. These projects have almost all followed the dominant paradigm — launched 45 years ago by Cocconi and Morrison in the pages of this journal¹ — of using radio telescopes to look for signs of extraterrestrial life. This focus on electromagnetic waves (primarily at radio wavelengths, but also at optical ones) was based on various arguments for their efficiency as a means of interstellar communication. On page 47 of this issue, however, Rose and Wright² make the case that, if speedy delivery is not required, long messages are in fact more efficiently sent in the form of material objects — effectively messages in a bottle. Although the suggestion itself is not new^{3,4}, it had never before been backed up by quantitative analysis.

Unless the messages are short or the extraterrestrials are nearby, this ‘write’ strategy requires less energy per bit of transmitted information than the ‘radiate’ strategy does. Cone-shaped beams of radiation necessarily grow in size as they travel outwards, meaning that the great majority of the energy is wasted, even if some of it hits the intended target. A package, on the other hand, is not ‘diluted’ as it travels across space (Fig. 1), presuming that it’s correctly aimed at its desired destination. For short messages, however, electromagnetic waves win out because of the overheads



As an example of a large message, consider all of the written and electronic information now existing on Earth: it's estimated⁵ to amount to about one exabyte (that's 10^{18} bytes, or 10^{19} bits). Rose and Wright calculate that, using scanning tunnelling microscopy, these bits could be inscribed (in nanometre squares) within one gram of material! But this precious package would still require a cocoon of 10,000 kilograms to accelerate it from our planet to a speed of 0.1% of the speed of light, protect it from radiation damage along a 10,000-light-year route, and then decelerate it upon arrival.

•	= 1	= 1	--	= 12	
••	= -	= 2	---	= 24	
•••	=	= 3	--	= 100	= 10^2
••••	= --	= 4		= 1000	= 10^3
•••••	= -	= 5	2+3	= 5	
••••••	= --	= 6	8+17	= 25	$5 + \frac{2}{3} = 5\frac{2}{3}$
		= 7	$\frac{1}{2} + \frac{1}{3} = \frac{5}{6}$		$2 \times 3 = 6$
	---	= 8	$\frac{1}{3} + \frac{1}{5} = \frac{8}{15}$		$13 \times 28 = 364$
	--	= 9			
	--	= 10			

bit, as these authors assume. Furthermore, we do not know if such packages, even if efficiently sent, would ever in fact be recognized and opened. Of course, we also do not know whether electromagnetic signals intended for us — and which may in fact be now bathing the Earth — will ever be recognized as such. In both cases, repetition and diversity of communications media would seem to increase the extraterrestrials' and our chances of success.

So how should these results influence today's SETI strategy? Short "we are here" messages would still seem to be most efficiently sent by electromagnetic waves, and we should continue looking for the same. But perhaps some attention should be paid to the possibility of one day finding in our Solar System an information-drenched artefact, sent by an extremely advanced extraterrestrial civilization interested only in one-way communication. This intruder might be orbiting the Sun or a planet, or resting somewhere on a planet, moon or asteroid. The scenario is reminiscent of Arthur C. Clarke's *2001: A Space Odyssey*, in which a monolith discovered on the Moon has been left by extraterrestrials. If

astroarchaeologists were to find such an object, it would hardly be the first time that science fiction had become science fact. ■

Woodruff T. Sullivan III is in the Department of Astronomy, University of Washington, Seattle, Washington 98195, USA.

e-mail: woody@astro.washington.edu

1. Cocconi, G. & Morrison, P. *Nature* **184**, 844–846 (1959).
2. Rose, C. & Wright, G. *Nature* **431**, 47–49 (2004).
3. Bracewell, R. *Nature* **187**, 670–671 (1960).
4. Papagiannis, M. Q. *J. R. Astron. Soc.* **19**, 277–281 (1978).
5. Murphy, C. *Atlantic* **277**, No. 5, 20–22 (1996).

Evolutionary biology

Time, space and genomes

Nipam H. Patel

In most animals, the Hox genes — which control development — are clustered together. But why? New evidence supports the idea that the requirement for a temporal order of expression keeps the cluster intact.

Some of the most striking discoveries in developmental biology over the past century concern the set of genes called homeotic (Hox) genes. Genetic studies in fruitflies first showed that these genes have a major role in producing the head-to-tail (anterior–posterior) pattern of tissues along the body axis. Then came the startling finding that the order of these genes along a chromosome correlates with the anterior–posterior position of the body regions they control, and with the domains in which the genes are expressed. It soon became apparent that the same relationship exists in other animal groups, including vertebrates. Intriguingly, however, it seems that somewhere in the evolutionary lineage leading to the tunicate *Oikopleura dioica*, the Hox complex has disintegrated, as Seo and colleagues report on page 67 of this issue¹.

Evolutionary analyses have suggested that the common ancestor of bilaterally symmetrical creatures — which include most animals, the main exceptions being cnidarians and sponges — probably possessed at least seven Hox genes, organized into a single complex. Within the lineage leading to vertebrates, gene duplications led to an expansion of the cluster, and then the cluster itself underwent duplications, leading to the four copies of the Hox complex now found in humans and mice. All along, the collinearity of the genes — the correspondence between their physical order along chromosomes and their domains of expression and function — was maintained.

But why has collinearity been preserved? The ancestral bilaterian complex itself probably arose through several rounds of local duplications, explaining how the genes first became organized as a cluster. In general, however, gene order is constantly shuffled by chromosomal rearrangements such as inversions and movements of large DNA segments. Given the rate at which this process occurs, the maintenance of collinear organization over at least 600 million years of evolution must not just be due to chance². One possibility is that different Hox genes

once shared, and continue to share, regulatory elements. But although this idea might account for the preservation of some degree of organization, it seems inadequate to explain the extent to which the complex has been maintained. Another possibility is that the mechanism that allows the genes to be expressed in a strict anterior–posterior expression pattern requires some type of higher-level organization, involving the progressive chemical or structural modification of a large contiguous stretch of DNA.

The work of Duboule and colleagues over the past few years has added an extra dimension to the issue of collinearity. They have shown that the vertebrate Hox genes show not just spatial but also temporal collinearity³; that is, genes at one end of the complex are expressed not only in the anterior of the embryo, but also relatively early in development. Hox genes located further along the complex are expressed both more posteriorly and later. Duboule and colleagues⁴ have provided evidence that it may be the requirement to maintain temporal collinearity that is responsible for keeping the complex together. A Hox gene experimentally moved around within the complex can retain spatial information, but will have an altered temporal expression profile.

Continuing this theme, Seo *et al.*¹ provide

a fascinating example of an animal in which the Hox complex has not stayed together yet appears to maintain some degree of ordered spatial expression along the anterior–posterior axis. Their studies focus on *Oikopleura dioica* (Fig. 1). *Oikopleura* is a type of tunicate, but is quite distant from *Ciona*, the other well-studied representative of this group of animals. Tunicates are evolutionarily primitive relatives of vertebrates, and comparisons between living tunicates and vertebrates may help researchers to piece together the features of the common invertebrate ancestor that gave rise to vertebrates. *Oikopleura* also has a remarkable genome — it is very small (at 60–70 megabases) and compact (with one gene every 4 kilobases)⁵.

Seo *et al.* find that *Oikopleura* has a complement of nine Hox genes. As expected, *Oikopleura* counterparts of the vertebrate anterior Hox genes are expressed in anterior regions of the developing animal, and counterparts of progressively more posterior vertebrate Hox genes are expressed in correspondingly more posterior regions. What is fascinating, however, is that the *Oikopleura* Hox genes retain this pattern of expression even though they are no longer in any sort of complex. Seo *et al.* show that for at least eight of these genes, no other Hox gene is found within 250 kilobases on either side. It is not that these Hox genes are in a gene-poor region, however; each is surrounded by other genes at the usual high density found in this animal.

These results, then, would seem to indicate that spatial collinearity can be maintained without requiring the organization of the Hox genes into a complex. *Oikopleura*, however, appears not to maintain temporal collinearity. The expression of its Hox genes does not seem to begin in a progressive temporal order, but rather at roughly the same time. Extensive splits within the Hox complex are also seen in the roundworm *Caenorhabditis elegans* and in *Ciona*, two other cases in which temporal collinearity no longer applies. Even in fruitflies, the Hox complex is split into two, and many non-Hox

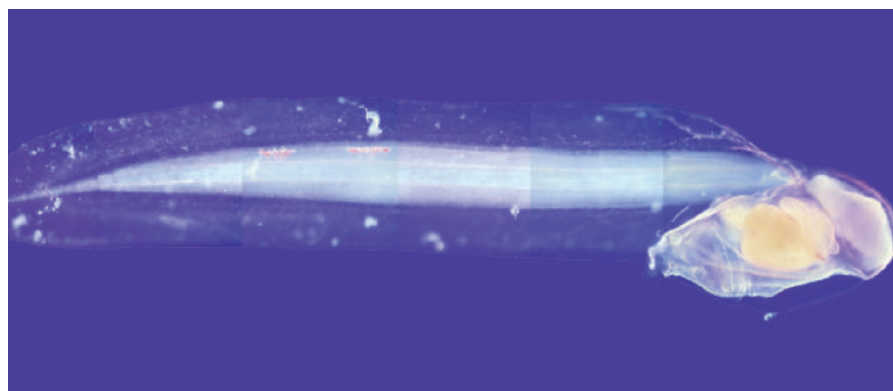


Figure 1 Distant relative. *Oikopleura dioica* has a simple body plan reminiscent of that of a tadpole, hinting at its close affinity to vertebrates. Remarkably, the generation time for this organism is only four days.

R. RUDOLF

genes sit within the clusters. Although studies of fruitflies established the rules of spatial collinearity, this insect shows little, if any, evidence of temporal collinearity. On the other hand, evolutionarily more primitive insects appear to maintain temporal collinearity in Hox-gene expression. These findings tend to support the idea that temporal collinearity does require an intact Hox complex — although this may not be the only selective force holding the cluster together.

Another interesting lesson is emerging from comparisons of fruitflies, *C. elegans*, *Ciona* and *Oikopleura* with animals such as humans and mice. The former group of species make excellent experimental models, because of their rapid development and small genomes. But it is curious that the presumed ancestral bilaterian Hox complex has

been split to varying degrees in each of these model systems — and temporal collinearity no longer applies. It might be that other aspects of their development and genome organization also differ substantially from their slower-growing and relatively large-genomed relatives, including humans. To be fair, however, the conclusion will probably be that every animal has its own unique and fascinating properties. ■

Nipam H. Patel is in the Departments of Molecular and Cell Biology, and of Integrative Biology, University of California at Berkeley, Berkeley, California 94720-3140, USA.
e-mail: nipam@uclink.berkeley.edu

1. Seo, H.-C. *et al.* *Nature* **431**, 67–71 (2004).
2. Patel, N. H. & Prince, V. E. *Genome Biol.* **1**, reviews1027 (2000).
3. Kmita, M. & Duboule, D. *Science* **301**, 331–333 (2003).
4. van der Hoeven, F., Zakany, J. & Duboule, D. *Cell* **85**, 1025–1035 (1996).
5. Seo, H.-C. *et al.* *Science* **294**, 2506 (2001).

Statistical physics

Hear the noise

Simon Kos and Peter Littlewood

At the nanoscale, thermal fluctuations and noise dominate. But instead of being a hindrance, the details of the noise itself can reveal the physical properties of the system.

Almost two centuries ago, the atomic nature of matter was elegantly revealed by brownian motion — as exemplified by the random motion of pollen particles in water as they are bombarded by water molecules. Now Crooker *et al.*¹ provide, on page 49 of this issue, a textbook demonstration of fluctuations at work in the spectroscopy of a small number of alkali atoms — a quantum version of brownian motion.

In 1905, Einstein² pointed out a subtle consequence of the fluctuations in classical brownian motion: the same random forces that make a pollen particle jitter would also cause friction if the particle were dragged through the water. In other words, the fluctuation of the particle at rest has the same origin as the dissipation of the motion of a moving particle that is subject to an external force. Einstein's result is a general one, codified in the 'fluctuation–dissipation theorem', which is one of the deepest results of thermodynamics and statistical physics. Einstein's observation also had a crucial consequence for the state of thermal equilibrium, in which the fluctuation of the stationary particle is characterized by a single number (the diffusion constant), and the friction of a moving particle is characterized by another number (the mobility). The fluctuation–dissipation theorem states that these two numbers, previously considered independent of each other, are, in fact, connected by a simple relation: they are proportional to each other

through the absolute temperature, a characteristic of the state of thermal equilibrium.

Brownian motion follows Newton's classical equations. However, the invention of quantum mechanics, and the discovery of Heisenberg's uncertainty principle, brought discrete energy levels and new kinds of fluctuation into play. It took several decades to develop the proper techniques of quantum statistical physics, and the fluctuation–dissipation theorem did not reach its final form until the work of Kubo in the 1950s. Its formal derivation for quantum systems is even more mathematical and less transparent than in the classical case; many of us struggled to understand the physical context of this elegant but esoteric mathematical statement when we first met it in an undergraduate course. But the final result again has a clear physical interpretation: the dissipation in a quantum system is caused by transitions between its energy levels; the noise spectrum should, therefore, have peaks at frequencies corresponding to the differences in energy between these levels. All of the physical properties of a quantum system depend on the values of these energy levels: the 'colour' of an atom and its nuclear magnetic resonance frequency are two examples that involve measuring the energy spectrum.

Quantum physics began with the study of the emission lines of the hydrogen atom — the particular wavelengths of radiation emitted as the atom's electron makes transitions between energy states. In their experiment,

Crooker *et al.*¹ study the magnetic fluctuations in vapours of hydrogen-like alkali atoms of rubidium and potassium. These atoms have a single valence electron, whose direction of spin, or magnetization, can change. Crooker *et al.* measure birefringence, a phenomenon well known in solutions of helical organic molecules such as sugars: because of the helicity of the molecules, right- and left-polarized light propagates through the solution with different speeds, and hence incident linearly polarized light leaves the sample with its direction of polarization rotated.

The same effect happens for alkali atoms in their ground state where the helical orientation is provided by the spin of the valence electrons. Crooker and colleagues' experiments show that the polarization of light fluctuates in time, as the spin magnetization itself fluctuates. But rather than following the maxwellian distribution of classical thermal noise, the temporal fluctuations have a complex spectral structure — hyperfine structure — owing to a delicate interaction between the spin of the valence electron and that of the atomic nucleus. These lines can be resolved, as predicted, despite being narrower than the linewidth of the probe laser. Noise produced by a summation of random events grows more slowly than the system size. Crooker *et al.*¹ show that the noise per atom scales inversely as the square root of the number of atoms, as expected.

Often we study systems by perturbing them — by measuring their response to an external probe. But this approach becomes increasingly difficult for the small systems that are now the focus of many studies in nanoscale or biological sciences. These experiments¹ remind us that 'listening' to the intrinsic noise of a system in equilibrium can provide the same information as does probing it with an external field (which in the present case would be equivalent to performing a conventional magnetic resonance experiment on the electron spins). Crooker *et al.* have provided an elegant example of a general principle, one that might be exploited, for example, in chemical sensors by measuring the thermal vibration of small cantilevers³, or could be used to measure the fluctuations of a single electron spin on a surface using a scanning tunnelling microscope^{4,5}. For the moment, however, it is a practical demonstration of an arcane yet fundamental piece of science, first intuited by Einstein during his *annus mirabilis* a century ago. ■

Simon Kos and Peter Littlewood are in the Theory of Condensed Matter Group, Cavendish Laboratory, Madingley Road, University of Cambridge, Cambridge CB3 0HE, UK.
e-mail: pbl21@cam.ac.uk

1. Crooker, S. A., Rickel, D. G., Balatsky, A. V. & Smith, D. L. *Nature* **431**, 49–52 (2004).
2. Einstein, A. *Ann. Phys.* **17**, 549–560 (1905).
3. Paul, M. R. & Cross, M. C. *Phys. Rev. Lett.* **92**, 235501 (2004).
4. Balatsky, A. V. *et al.* *Phys. Rev. B* **66**, 195416 (2002).
5. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. *Nature* **430**, 329–332 (2004).

Cell biology

Regulated self-cannibalism

Daniel J. Klionsky

Cells consume parts of themselves to survive starvation and during development. But how do they control this process of self-eating so that it begins at the right time and does not end up killing the cell?

When you are fasting or are otherwise deprived of food, your body starts to break down stored nutrients to keep essential processes and organs (such as your brain) supplied with fuel. Similarly, when a cell is deprived of nutrients it will degrade some of its own constituents to stay alive. It does this by the process of autophagy — literally, 'self-eating'. Writing in *Developmental Cell*, Scott and colleagues¹ and Rusten and co-workers² provide new insights into how autophagy is regulated in the cell, both as a response to starvation and during development.

Autophagy is marked by the formation of autophagosomes — large vesicles, bounded by a double membrane, that sequester cytosol and organelles such as mitochondria. Fusion of an autophagosome with a lysosome releases the inner vesicle into the lysosome, where enzymes break the vesicle down. Degradation and recycling of the vesicle's constituents enable the cell to continue to carry out essential processes.

Because autophagy has the capacity to degrade entire organelles, it could be harmful if it occurred randomly. Autophagy must therefore be tightly regulated. The question is: how is this regulation accomplished? How does a cell or organism sense its environment and trigger an appropriate signal to induce or suppress autophagy? The process occurs in organisms ranging from yeast to worms to humans³, and Scott *et al.*¹ and Rusten *et al.*² chose the fruitfly *Drosophila melanogaster* as their subject.

Drosophila larvae have a storage organ called the fat body, which has analogous functions to those of the liver and adipose (fat) tissue in vertebrates: it is a source of large amounts of potential energy. If an animal has just eaten, of course, there is no need to tap into these energy stores. In response to food, the hormone insulin is produced; this binds a receptor on the surface of cells in the fat body and throughout the organism, and triggers a signalling cascade (Fig. 1). An important part of this cascade is an enzyme termed class I

phosphatidylinositol-3-OH kinase (PI(3)K). This adds a phosphate group to a particular position on the lipid phosphatidylinositol, which is part of the cell membrane. Various proteins in turn bind the phosphorylated lipid and become activated, thus transmitting the external signal (in this case, insulin) into the cell. A central player in this pathway is the enzyme Tor, which is involved in many regulatory events connected with energy metabolism, and which suppresses autophagy⁴.

So this pathway provides a mechanism by which cells can block self-eating if the organism has just fed. On the flip side, the pathway also provides a means of inducing autophagy when the organism is starved: a lack of food (particularly of the sugar glucose) leads to a lack of insulin, leaving the pathway inactive and enabling the cell to tap into its storage reserves. But how does the cell regulate the degree of autophagy so that it does not get out of hand?

An interesting explanation for this additional level of control comes from Scott and colleagues¹. The activity of the enzyme p70 S6 kinase has been shown increase when Tor is turned on — that is, when nutrients are abundant^{5,6}. This led to the proposal that p70 S6 kinase is itself a negative regulator of autophagy⁷. However, Scott *et al.* now show that this enzyme must be active for autophagy to be maximally induced.

How can this finding be reconciled with the fact that Tor activates p70 S6 kinase? The result seems to suggest that p70 S6 kinase needs to be active under conditions (starvation) in which its activator is turned off. Perhaps, after Tor is switched off, any active p70 S6 kinase remains active for some time, allowing maximal autophagy. But other mechanisms might then gradually deactivate the p70 S6 kinase, thereby preventing excessive autophagy, which could be harmful. These authors also show that autophagy is essential in supplying nutrients to promote survival in the absence of Tor function. Thus, a starvation signal results in the down-regulation of insulin/PI(3)K signalling, followed by a consequent inhibition of cell growth, and a concomitant upregulation of autophagy to supply essential nutrients.

In simple organisms such as yeasts, autophagy is primarily a response to starvation. But in more advanced organisms such as *Drosophila*, autophagy is also involved in various developmental pathways⁸. This adds another level of complexity to its regulation, because there are more occasions when an organism may need to activate autophagy — during growth, for example, or, in an insect, when the pupa forms. A starvation response must be able to cope with the unexpected (a sudden loss of food), but for the purposes of development, cellular responses, including autophagy, must be programmed. In other words, there must be a mechanism to initiate autophagy at precise times. This is achieved

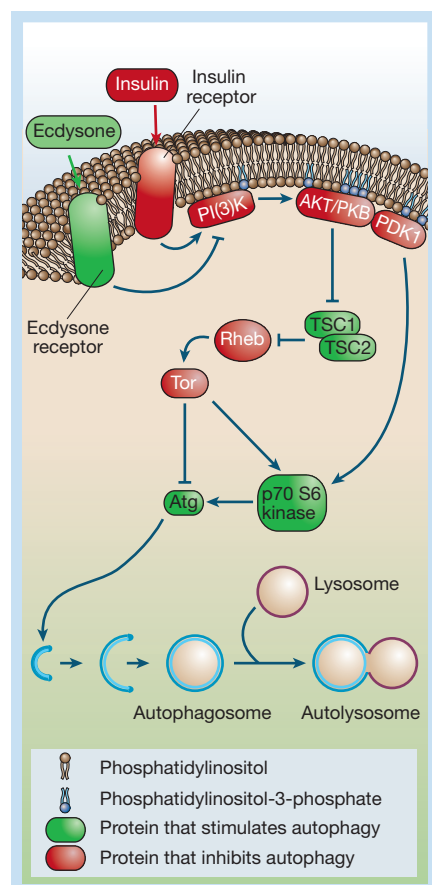


Figure 1 Cellular self-eating: the basics. Insulin produced in response to food inhibits autophagy. This signal is passed into the cell by way of the insulin receptor, which activates the enzyme phosphatidylinositol-3-OH kinase (PI(3)K). This in turn modifies the lipid phosphatidylinositol in the cell membrane, leading to the activation of other enzymes, including AKT/PKB and PDK1. AKT/PKB inhibits TSC1–TSC2, which can therefore no longer inhibit Rheb; Rheb then activates Tor, which suppresses autophagy by inhibiting Atg proteins. Rusten *et al.*² suggest that ecdysone, a hormone in fruitfly larvae, overrides the nutrient-inhibited response and activates autophagy by inhibiting PI(3)K, although the exact mechanism is unknown. Scott *et al.*¹ show that p70 S6 kinase is needed for full activation of autophagy, and may prevent excessive autophagy when Tor is inactive. Autophagy involves the formation of double-membraned vesicles that sequester cytoplasm and organelles. Fusion of these autophagosomes with lysosomes containing acid hydrolase enzymes allows the degradation and recycling of macromolecular constituents that are needed for cell survival under starvation conditions and for remodelling of the organism during development.

Theoretical biology

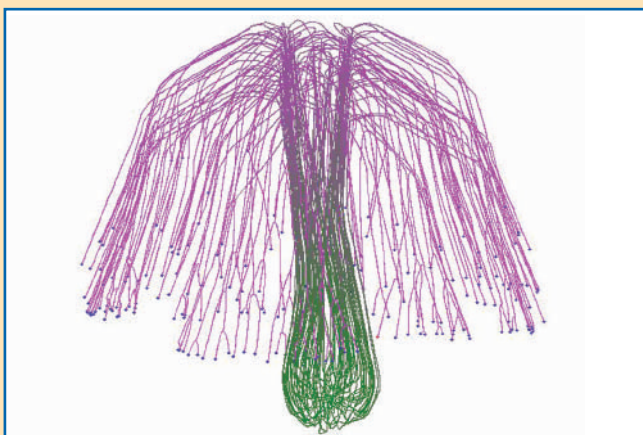
Mushrooms in cyberspace

Mushrooms arise largely from the inflation of pre-existing cells, which in part accounts for the startling speed with which they can appear. But what developmental processes are responsible for shaping those cells first into the primordial mushroom structure and then into the full-grown 'fruiting body' itself? Audrius Meškauskas and colleagues, writing in *Mycological Research* (108, 341–353; 2004), provide a new angle on this question. They have grown cyber fungi like the mushroom shown here. As well as creating primordial fruiting bodies whose cell arrangements mimic the real things, the authors' computer models provide predictions that can be tested.

Fungi use a single cell type — the filamentous hypha — to generate mushrooms and other multicellular organs such as the cords and rhizomorphs that function as exploratory devices for colonies once they run out of food. This

reliance on hyphae distinguishes fungi from plants and animals, both of which produce a variety of cell types that are specialized for different functions. For this reason, a model of mushroom development need only specify the positions of cells. Unfortunately, simple anatomy has not led to a clear explanation of the processes that make cells lying parallel to one another in the stem of a mushroom, blossom into the bell-shape of the cap.

What Meškauskas *et al.* show is that baby cyber mushrooms develop simply by applying rules of mutual attraction and repulsion to every one of thousands of gravity-sensing hyphae. As long as all of the filaments behave in precisely the same way at the same time, there is no requirement for the exercise of global, or organ-level, control in fabricating the whole structure. This means that the intricate shapes of different mushrooms might be specified in a clockwork fashion,



solely by genes activating successive waves of cellular attraction and repulsion. This is good news for mycologists, because the kinds of hormone-pumping meristems that are found at the tips of shoots and roots, and that regulate plant development, have not been found in fungi.

With the evidence of the computer animations, experiments on single cells take on new significance. Manipulation of genes that steer fungi in rotting wood, for example, may be likely to change

the arrangement of cells in a mushroom and result in some extraordinary fruiting body forms. If more progress can be made in understanding what makes a hypha bend towards or away from its neighbours, mycologists will be closer to solving the mystery of mushroom development.

Nicholas P. Money

Nicholas P. Money is in the Department of Botany, Miami University, Oxford, Ohio 45056, USA.

e-mail: moneynp@muohio.edu

through the use of specific hormones. In *Drosophila*, the main hormone that controls pupation is ecdysone.

Rusten *et al.*² focus on the developmental regulation of autophagy in the *Drosophila* fat body, where cells undergo developmentally programmed autophagy during the larval stage. In contrast to the inhibitory role of PI(3)K signalling, the hormone ecdysone appears to promote autophagy. But how are these two regulatory factors coordinated during development? The authors find that the occurrence of autophagy increases from early through to late larval stages. This progression correlates with a gradual increase in the level of ecdysone and a corresponding decrease in the level of PI(3)K-modified lipids. Conversely, inactivation of ecdysone signalling has the opposite effect. Thus, ecdysone appears to negatively regulate PI(3)K signalling, allowing autophagy to occur (Fig. 1).

The two stories come together in the finding of Rusten *et al.* that developmentally programmed autophagy is subject to regulation by Tor: inhibiting this protein results in an increase in autophagy. The implication is that under normal conditions Tor is not completely inhibited during developmental autophagy in the fat body — some Tor molecules must be active, providing a pool that can be inhibited to increase the occurrence or amount of autophagy. The fat body therefore allows developmentally and

nutritionally triggered autophagy to be coordinated. An understanding of the nuances of this coordination — and whether it occurs in other organisms, including ourselves — must await further studies.

Daniel J. Klionsky is in the Departments of Molecular, Cellular and Developmental Biology, and of Biological Chemistry, Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109, USA.
e-mail: klionsky@umich.edu

1. Scott, R. C., Schuldiner, O. & Neufeld, T. P. *Dev. Cell* 7, 167–178 (2004).
2. Rusten, T. E. *et al.* *Dev. Cell* 7, 179–192 (2004).
3. Reggiori, F. & Klionsky, D. J. *Eukaryot. Cell* 1, 11–21 (2002).
4. Jacinto, E. & Hall, M. N. *Nature Rev. Mol. Cell Biol.* 4, 117–126 (2003) (Erratum: 4, 249 (2003)).
5. Chung, J., Kuo, C. J., Crabtree, G. R. & Blenis, J. *Cell* 69, 1227–1236 (1992).
6. Price, D. J., Grove, J. R., Calvo, V., Avruch, J. & Bierer, B. E. *Science* 257, 973–977 (1992).
7. Blommaert, E. F., Luiken, J. J., Blommaert, P. J., van Woerkom, G. M. & Meijer, A. J. *J. Biol. Chem.* 270, 2320–2326 (1995).
8. Levine, B. & Klionsky, D. J. *Dev. Cell* 6, 463–477 (2004).

Planet formation

The core problem

William B. Hubbard

Controversy over shock-wave experiments on the compression of hydrogen has broad implications — for understanding the cores of Jupiter and Saturn, and even the formation of extrasolar planets.

In July of this year, NASA announced that it will fund further study of a proposed mission to Jupiter, as part of its New Frontiers programme¹. A prime objective of this orbiter mission, called Juno, would be to measure Jupiter's gravitational and magnetic fields at very close range, to discern whether the planet has a dense core. Meanwhile, in the *Astrophysical Journal* Saumon and Guillot² write that the existence of a massive core in Jupiter may

depend on the validity of an experiment here on Earth.

Over the past decade, it has become possible in the laboratory to squeeze hydrogen to pressures more than a million times greater than atmospheric pressure, while simultaneously heating it to temperatures exceeding 10,000 kelvin. Compression experiments on the hydrogen isotope deuterium — driven by NOVA, a huge laser-implosion device at the US Lawrence Livermore National

Laboratory — seem to show that, under such conditions, deuterium's density increases as much as sixfold³. Other experiments, however, find that this hydrogen isotope is less compressible⁴. Theoretical calculations have yet to weigh in fully to help sort out the discrepancy. But Saumon and Guillot² find that if the NOVA results are right, then their models suggest that Jupiter does indeed have a sizeable core. The implications extend beyond the planned NASA mission to the very definition of what constitutes a planet.

Although Jupiter is 300 times more massive than Earth, it has only one-thousandth of the mass of the Sun; Saturn comprises about 100 Earth masses and is the only other hydrogen-rich planet in our Solar System. The formation of either planet by the spontaneous gravitational collapse of such a small mass of hydrogen from the Sun's primordial nebula — a swirling disk of gas and dust — has long been considered unlikely, especially because temperatures in the nebula could never be low enough to condense either hydrogen or helium. Furthermore, the Galileo probe spacecraft, which entered Jupiter's atmosphere in 1995, found that the outermost layers (at least) of Jupiter have a 'metallicity' about three times that of the Sun. To astrophysicists, the metallicity of a star refers to its enrichment in elements heavier than hydrogen and helium, relative to the Sun. 'Metals' comprise a bit more than 1% of the Sun's mass. The high metallicity of the jovian atmosphere would not be expected if the planet were formed by direct collapse from a solar-metallicity nebula⁵.

It is more likely that the formation of Jupiter and Saturn was initiated by the collapse of primordial nebular gas into the gravitational well created by a solid, dense core of refractory (readily condensed) elements (Fig. 1a). Estimates vary of the minimum mass of core needed to trigger the collapse, but typical values are around 15 Earth masses⁶. To put the problem in perspective, that mass of rock and ice would correspond to about 1,000 Earth masses of hydrogen and helium in a solar-metallicity nebula. About a third of that would need to be captured to make a Jupiter, and another tenth to make a Saturn.

Saturn's core stands out clearly in models of its planetary interior, and there is little doubt that it is about the right mass². But the canonical 15 Earth masses needed to trigger collapse is equivalent to only about 5% of Jupiter's mass. So to confidently predict the existence of a core, the compression curve of Jupiter's dominant component, hydrogen, must be known to much better than 5%. Using the 'softer' NOVA results³ for hydrogen compression, Saumon and Guillot² find that models of Jupiter with sizeable cores are possible; the core mass shrinks to less than 5 Earth masses (and possibly even to zero) when using the 'stiffer' shock data from the Z

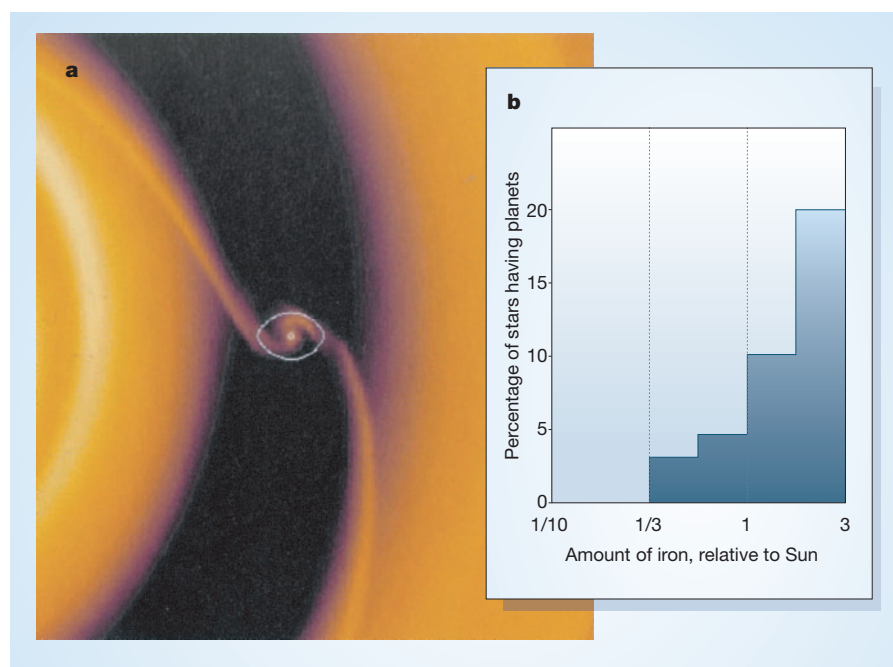


Figure 1 Formation of Jupiter and Jupiter-like planets. Using data³ on the compression of hydrogen, Saumon and Guillot² suggest that such planets are likely to have a dense core. a, Jupiter may then have formed from gas sucked in from the solar nebula onto its protoplanetary core (derived from ref. 10). b, Jupiter-like planets have been discovered orbiting stars outside the Solar System; the planetary population around a star seems to depend on the star's metallicity — the amount of elements heavier than helium that the star contains (here, iron is taken as the indicator of metallicity⁸). The higher a star's metallicity, the more likely it is that sizeable planetary cores would form — tying in with Saumon and Guillot's expectation of a dense core in such planets. However, Saumon and Guillot's models do not predict a dense core for Jupiter when they include conflicting experimental data⁴ that suggest that hydrogen is in fact much less compressible.

accelerator at Sandia National Laboratory⁴. Physicists might eventually resolve the discrepancy with more shock experiments, and first-principles theoretical simulations of dense hydrogen should also provide guidance.

A definitive answer on the formation of Jupiter has relevance beyond our Solar System. Well over 100 planets that are similar in mass to Jupiter or Saturn have now been detected in orbit around other stars. For a few of them, their radii have been measured and their deduced mean densities are similar to those of Jupiter and Saturn — that is, similar to that of water. Nothing but hydrogen could give such low mean density in a Jupiter-mass planet⁷. It has also been shown that the higher a star's metallicity, the more Jupiter-like planets are likely to orbit it⁸ (Fig. 1b). The seeming paucity of Jupiters orbiting low-metallicity stars is as expected if cores of the necessary size cannot form around such stars. The story would hang together nicely, if only we could be sure that Jupiter does have the core predicted.

The International Astronomical Union's Working Group on Extrasolar Planets has provisionally defined⁹ 'planets' to be 'objects with true masses below the limiting mass for thermonuclear fusion of deuterium (currently calculated to be 13 Jupiter masses for objects of solar metallicity) that orbit stars

or stellar remnants... no matter how they formed". No one is entirely happy with this definition, although it does jibe with the definition of the boundary between stars and brown dwarfs at roughly 80 Jupiter masses, where fusion of ordinary hydrogen (protons) ceases. If it could be shown that formation of Jupiter-like planets requires a dense core of at least 15 Earth masses, and that Jupiter itself has such a core, perhaps the definition of a planet could instead be based on the core-nucleation model, as suggested by astrophysical evidence. ■

William B. Hubbard is at the Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721-0092, USA.

e-mail: hubbard@lpl.arizona.edu

1. www.nasa.gov/lb/home/hqnews/2004/jul/HQ_04228_new_frontiers.html
2. Saumon, D. & Guillot, T. *Astrophys. J.* **609**, 1170–1180 (2004).
3. Collins, G. W. et al. *Science* **281**, 1178–1181 (1998).
4. Knudson, M. D., Hanson, D. L., Bailey, J. E., Hall, C. A. & Asay, J. R. *Phys. Rev. Lett.* **87**, 225501 (2001).
5. Hubbard, W. B., Burrows, A. & Lunine, J. I. *Annu. Rev. Astron. Astrophys.* **40**, 103–136 (2002).
6. Podolak, M., Hubbard, W. B. & Pollack, J. B. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 1109–1147 (Univ. Arizona Press, Tucson, 1993).
7. Burrows, A., Hubeny, I., Hubbard, W. B., Sudarsky, D. & Fortney, J. J. *Astrophys. J.* **610**, L53–L56 (2004).
8. Fischer, D., Valenti, J. A. & Marcy, G. *IAU S219, ASP Conference Series* (in the press); http://exoplanets.org/iau_proc.pdf
9. www.ciw.edu/boss/IAU/div3/wgesp/definition.html
10. Lubow, S. H. & Artymowicz, P. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 731–755 (Univ. Arizona Press, Tucson, 2000).

Cancer

Cell survival guide

Eric R. Fearon and Kathleen R. Cho

A jaded observer might consider the cancer research field near maturity and surprising new results improbable. But work on the protein netrin-1 shows that unforeseen insights into cancer can still occur.

Named from the Sanskrit word for 'one who guides', the netrin-1 protein was discovered because of its ability to direct the migration of axons in the developing spinal cord^{1,2}. The functions of netrin proteins in axon guidance and neural cell migration have consequently received much attention (reviewed in ref. 3). But the report on page 80 of this issue from Mazelin *et al.*⁴ makes the case that netrin-1 might offer unanticipated guidance for the cancer field as well.

Netrins are secreted proteins that act on neural cells through transmembrane receptors of the DCC and UNC5H families (reviewed in ref. 3). DCC stands for 'deleted in colorectal carcinomas', and as its name implies, DCC was initially described as a candidate tumour-suppressor gene. In colorectal cancer, a chromosomal region containing DCC is frequently deleted. This is often accompanied by low or absent expression of the DCC protein and, occasionally, by specific DCC mutations^{3,5,6} — as expected for a tumour-suppressor gene. However, it has been argued that DCC might not actually be a tumour-suppressor gene, but be functioning only in the nervous system^{7,8}. This point

of view is based on the fact that the mechanisms underlying reduced or absent DCC expression in most cancers are poorly defined^{3,5,6}, and the observation that mice lacking one copy of *Dcc* do not have an obvious predisposition to tumours⁷.

The three *UNC5H* genes — *UNC5H1*, *UNC5H2* and *UNC5H3* — are also proposed to be tumour-suppressor genes⁹. But concerns similar to those for DCC could be raised about the evidence that links defects in the *UNC5H* genes to cancer in humans. Possible mechanisms contributing to the loss of *UNC5H* gene expression in cancer include deletion or mutation of *UNC5H* genes, and transcriptional mechanisms⁹. For instance, the *UNC5H2* gene is regulated by the p53 tumour-suppressor protein¹⁰, and p53 function is often defective in cancer.

Perhaps the strongest suggestion that DCC and *UNC5H* are indeed involved in cancer comes from observations that, in cultured cells, the DCC and *UNC5H* proteins promote cell death when netrin-1 is absent but enhance cell survival when netrin-1 is present^{3,11,12}. This behaviour is consistent with the hypothesis that DCC and *UNC5H* belong to a class of cellular receptor called

dependence receptors, which induce programmed cell death unless they are occupied by their ligand. Theoretically, there are two mechanisms by which such receptors might cause cancer: either the ligand is present in excess or in the wrong location, so that inappropriate cell survival is encouraged; or the receptors are inactivated or deleted somehow so that cell death is no longer promoted when there is little or no ligand. The specific mechanisms by which netrin-1 binding to DCC or *UNC5H* affects cell survival or death remain unknown, but Mazelin and colleagues⁴ set out to test the dependence-receptor hypothesis *in vivo*.

They genetically engineered mice to express netrin-1 in epithelial cells in those parts of the intestinal lining where it is not normally expressed. As expected, they saw a reduction in cell death in small intestine and colon tissues. More notably, they found that these mice tended to develop spontaneous pre-cancerous intestinal epithelial growths (adenomas), with 17% of netrin-1 transgenic mice developing at least one adenoma, whereas control mice had none. Mice carrying constitutive mutations in the *Apc* (*adenomatous polyposis coli*) gene are well known to develop intestinal adenomas. Thus, to test the tumour-promoting effects of netrin-1 further, Mazelin *et al.* studied mice carrying the *Apc* gene defect together with the netrin-1 transgene. Compared with mice having just the *Apc* defect, mice carrying both the transgene and the defect had many more 'high-grade' adenomas or lesions with markedly aberrant cellular and glandular morphology. What's more, about 50% of

Astronomy

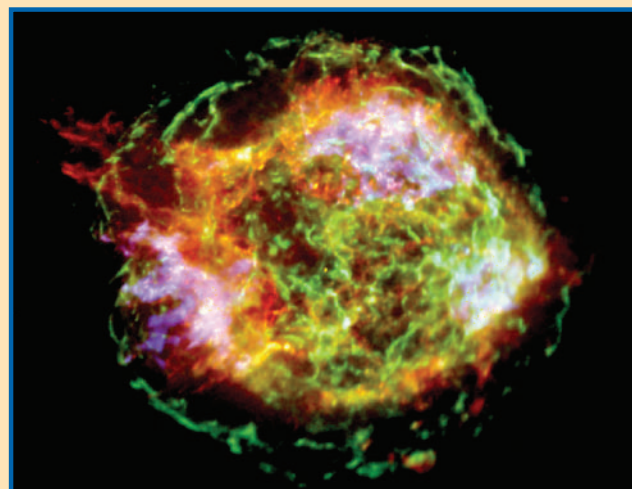
The quiet one

No one noticed the death of Cassiopeia A. But this star's supernova remnant, the youngest known in the Milky Way, has since become an icon of the success of the Chandra X-ray Observatory. Chandra produced its first images of Cassiopeia A in 1999, soon after its launch. Five years later, these images have become exquisite in their detail.

From the size and rate of expansion of the remnant, it is estimated that the light from Cassiopeia A's supernova explosion reached Earth more than 300 years ago. Over the course of history, firm sightings of other supernovae have been made, noted by contemporary astronomers as 'new stars' in the sky. Cassiopeia A, it seems, might have been spotted only by John Flamsteed, Britain's first

Astronomer Royal. In 1680, he recorded a star close to Cassiopeia A's position — but subsequent star catalogues have tended to regard this as an error. Anyway, an explosion date in the 1660s would fit with the observed expansion better.

In all likelihood, Cassiopeia A just wasn't very bright on the supernova scale of things. But at radio wavelengths, it is now the brightest object in the sky. And in the X-ray emission picked up by Chandra, details of Cassiopeia A's structure and composition are clear. The shock wave created by the supernova (the outer green ring in this picture) now spans a diameter of 10 light years; bipolar jets of mostly silicon atoms (one of which can be seen, in red, on the upper left) suggest that such a feature might be more



common in supernovae than had been thought.

The tiny yellow speck at the centre of this 1-million-second exposure is a neutron star — an incredibly dense mass of collapsed stellar material formed from the supernova. Cassiopeia A's neutron star is strangely quiet: it is not

firing out pulses of radiation as other neutron stars do. The Chandra team suspects that this star might have a particularly strong magnetic field, even qualifying as a 'magnetar' — a class of neutron stars with magnetic fields a thousand trillion times stronger than Earth's.

Alison Wright

NASA/CXC/GSFC/U. HWANG ET AL.



100 YEARS AGO

The value and possibilities of wireless telegraphy as a journalistic adjunct are described in Saturday's *Times* by the special correspondent who established a wireless telegraph system at the theatre of war operations in the Far East with such success that both the belligerents regarded the enterprise as dangerous to their interests. The Japanese Government placed such limitations upon the free movements of the *Haimun* — the vessel chartered by the *Times* for its wireless telegraph service — that this means of communication was discontinued of necessity; and there seems little doubt that in future the use of all systems of wireless communication will be controlled by international law. From *Nature* 1 September 1904.

50 YEARS AGO

A New Periodic Table of the Elements Based on the Structure of the Atom. To demonstrate the periodicity in the properties of the chemical elements, Lothar Meyer chose the most direct representation: he plotted the elements in the order of their atomic weights on the abscissa, and the values of the property in question on the ordinate. At the same time, Mendeléeef published the Periodic Law in the form of tables... During the following decades innumerable attempts were made to improve on Meyer and Mendeléeef. All sorts of representations, trees with branches, concentric circles, spirals, figure-eights, and various three-dimensioned curves were tried;... behind all this was the hope to get nearer to the mystery of the periodic system if a more perfect arrangement could be found. But the scientific result of all these attempts was nil... It is somewhat astonishing to see that quite recently "A New Periodic Table of the Elements" has been published which is a revival of the old discarded attempts. The curves, for example, which represent the specific gravity of the elements, are based on one of the well-tried spirals and can, naturally, not avoid the old drawbacks... The author recommends even a cone-shaped periodic chart, another repetition of previous suggestions; whoever takes the trouble to follow the advice to cut the drawing out and to gum it together as a cone, will scarcely get any insight into the sequence of chemical elements which the usual tables do not give. From *Nature* 4 September 1954.

high-grade adenomas in mice carrying the transgene and the *Apc* defect contained foci of invasive carcinoma, whereas no carcinomas were seen in mice with only the *Apc* defect.

In the normal small intestine and colon, epithelial cells proliferate predominantly in the bottom of small pits called crypts, before becoming specialized and migrating upwards, where cells die and slough off (Fig. 1). In normal mice, Mazelin *et al.* found that *Dcc* was uniformly expressed throughout the intestinal epithelium, whereas netrin-1 expression was mostly restricted to the base of the crypts. This is consistent with the view that interaction of netrin-1 with its receptors might regulate survival versus cell death, with netrin-1 stimulating proliferation in the crypt and the absence of netrin-1 perhaps contributing to cell death at the surface.

With this paper, Mazelin *et al.* offer encouraging *in vivo* data to support the hypothesis that the DCC and UNC5H proteins function as dependence receptors for netrin-1. Yet there are some caveats. First, because there are other netrin-1 receptors^{13,14}, an excess of netrin-1 might promote cell survival in part through pathways independent of DCC and UNC5H. Second, in the genetically engineered mice there is a uniform gradient of netrin-1 in the intestinal lining, as well as overexpression of netrin-1 relative to its usual levels, which could lead to some doubts about the physiological relevance of the findings. However, Mazelin *et al.* did see increased cell death in intestinal crypts of mice deficient in netrin-1, implicating netrin-1 in the survival of normal intestinal cells.

Their findings are consistent with the notion that there might be strong selection for inhibition of DCC and UNC5H expression or function in colon and other cancers. The proteins might indeed function in tumour suppression, perhaps by inhibiting growth or causing the death of potential cancer cells in environments where the netrin-1 concentration is low. Conceivably, singular inactivation of *DCC* or *UNC5H* might be insufficient to promote tumour development. Concerted inactivation of both genes might be required for progressive outgrowth of cells in regions where netrin-1 levels are usually low, explaining why mice lacking one copy of the *Dcc* gene showed no overt predisposition to intestinal or other tumours⁷.

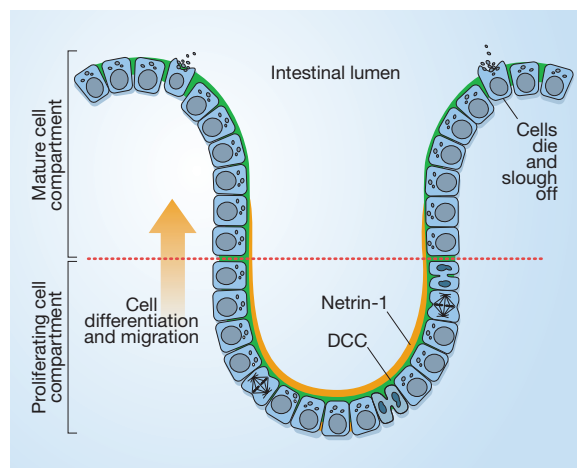


Figure 1 Model for the role of netrin-1 and DCC in regulating cell survival and death in the intestine, based on Mazelin and colleagues' results⁴. In a normal intestine, cells proliferate at the bottom of pits called crypts. They then migrate upwards towards the surface. As they go, the cells stop proliferating, become more specialized, and ultimately die. Netrin-1 expression is mostly restricted to the base of the crypt, but expression of its receptor DCC is essentially uniform in all epithelial cells⁴. According to the dependence-receptor hypothesis, the findings imply that binding of netrin-1 to DCC might contribute to cell survival in the crypt. At the surface, the absence of netrin-1 leads to DCC-mediated cell death.

To understand better the roles of DCC and UNC5H proteins as netrin-1 dependence receptors that regulate cell survival, it might prove useful to study other genetically engineered mice, such as mice with intestinal-specific deletions of the *Dcc* and/or *UNC5H* genes, or mice carrying *Dcc* or *UNC5H* mutations that are predicted to interfere with the receptors' ability to initiate cell death. Other dependence receptors have been identified³ and yet others probably remain to be discovered, so further research should help to clarify whether alterations in dependence-receptor pathways have a more widespread role in cancer.

Eric R. Fearon and Kathleen R. Cho are in the Division of Molecular Medicine and Genetics, Departments of Internal Medicine and Pathology, University of Michigan School of Medicine and Comprehensive Cancer Center, Ann Arbor, Michigan 48109-2216, USA.

e-mail: fearon@umich.edu

1. Serafini, T. *et al.* *Cell* **78**, 409–424 (1994).
2. Kennedy, T. E., Serafini, T., de la Torre, J. R. & Tessier-Lavigne, M. *Cell* **78**, 425–435 (1994).
3. Mehlen, P. & Mazelin, L. *Biol. Cell* **95**, 425–436 (2003).
4. Mazelin, L. *et al.* *Nature* **431**, 80–84 (2004).
5. Fearon, E. R. *et al.* *Science* **247**, 49–56 (1990).
6. Fearon, E. R. *Biochim. Biophys. Acta* **1288**, M17–M23 (1996).
7. Fazeli, A. *et al.* *Nature* **386**, 796–804 (1997).
8. White, R. L. *Cell* **92**, 591–592 (1998).
9. Thiebaut, K. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 4173–4178 (2003).
10. Tanikawa, C. *et al.* *Nature Cell Biol.* **5**, 216–223 (2003).
11. Mehlen, P. *et al.* *Nature* **395**, 801–804 (1998).
12. Llambi, F., Causeret, F., Bloch-Gallego, E. & Mehlen, P. *EMBO J.* **20**, 2715–2722 (2001).
13. Srinivasan, K., Strickland, P., Valdes, A., Shin, G. C. & Hinck, L. *Dev. Cell* **4**, 371–382 (2003).
14. Yebra, M. *et al.* *Dev. Cell* **5**, 695–707 (2003).

Signal transduction

Calcium lands lead role in cell cycle

J. Cell Biol. doi:10.1083/jcb.200402136 (2004)

Calcium ions have been implicated in many aspects of cell function, but a role in the cell cycle — the sequence of events by which cells replicate — has been speculative. Now, however, Violaine Sée *et al.* have found that Ca^{2+} ions can initiate the cell cycle.

The authors used Ca^{2+} -sensitive dyes to detect changes in Ca^{2+} levels inside connective-tissue cells called fibroblasts. When non-dividing cells were treated with growth-promoting serum, Ca^{2+} levels increased for around 30 seconds. This in turn triggered the activation of a regulatory factor called NF- κB — which is needed for entry into the cell cycle — and the expression of cyclin D1, a protein essential for progression through the cycle.

But Ca^{2+} ions alone were not sufficient to induce these effects: a family of enzymes known as mitogen-activated protein kinases (MAPKs) was also necessary. Sée *et al.* conclude that, in fibroblasts at least, the cell cycle is controlled by serum-dependent Ca^{2+} signalling through the MAPK–NF- κB pathway.

Helen Pilcher

Developmental biology

Given a hand

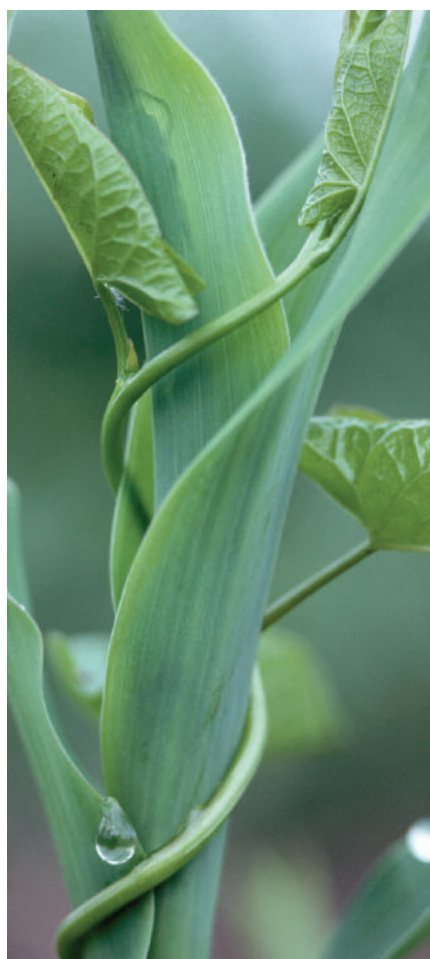
Cell **118**, 517–528 (2004)¹; *Cell* **118**, 505–516 (2004)²

In vertebrate embryos, the protein Sonic hedgehog (Shh) is produced in the zone of polarizing activity (ZPA) in the bud of tissue that will form a limb, and forms a gradient across the limb. It has been generally assumed that cells closer to the ZPA receive a higher dose of Shh than those further away and that this tells cells whether to form digit one, two, three, four or five in a hand or foot.

But matters might not be so simple. Brian D. Harfe *et al.*¹ have looked at cells in the mouse limb that start out making Shh in the ZPA, and found that they ended up in digits three, four and five, closest to the ZPA. They suggest that cells move out of the ZPA at different times and that, in this part of the limb, the time that cells spend in the ZPA and are exposed to Shh helps to determine which digit they will become.

Meanwhile, Sohyun Ahn and Alexandra L. Joyner² used the activity of the *Gli1* gene to detect cellular responsiveness to Shh. They found that many cells near the ZPA stop responding to Shh even while they are still exposed to it. So other unknown factors, which regulate the cells' reaction to Shh, might also help determine digit identity.

Helen Pearson



Evolution

Climb to success

Proc. R. Soc. Lond. B doi:10.1098/rspb.2004.2827 (2004)

'Key innovations' are adaptive traits that confer significant evolutionary success, such that groups that have adopted the adaptation are richer in species diversity than comparable groups that have not. A wide-ranging analysis of plant taxa now suggests that the climbing lifestyle of vines is one such trait.

Ernesto Gianoli scoured the taxonomy catalogues of 45 families of flowering plants, and came up with 48 groups of climbers, each matched to a sister group of non-climbers. In 38 cases, the climbers were more diverse. What's more, the pattern was evident when woody and non-woody pairs of plant groups were compared independently.

A climbing habit is a natural candidate for a key innovation. Provided that suitable support is available (in the form, for instance, of sturdy companion plants such as trees), climbers can reach the top of the canopy and achieve maximum exposure to sunlight. It is perhaps no surprise that climbing has evolved on many separate occasions during plant evolution, and that more than 130 plant families include vine-growing species.

Michael Hopkin

Optical imaging

Human lasers

Appl. Phys. Lett. **85**, 1289–1291 (2004)

Tumours can be spotted in human tissue by turning them into lasers. Randal C. Polson and Z. Vally Vardeny have found that human tissues infiltrated with organic dyes and illuminated by a standard neodymium:YAG laser may emit narrow-band light by a phenomenon called random lasing.

It has been known for more than a decade that light emitted within a collection of randomly positioned scatterers can give rise to lasing, characterized by sharp spectral lines. Under certain conditions this emission becomes coherent — not only is the light amplified and spectrally sharp, but the photons are in phase.

Polson and Vardeny induced coherent random lasing in a range of biological tissues, including potato and human colon tissue, stained with the dye Rhodamine 6G, which fluoresces with red light. They found more laser lines in the emission spectra from cancerous colon tissue than from healthy tissue, presumably because the diseased tissue was more disordered or more highly scattering (healthy tissue has a more uniform cell size). This allowed the researchers to distinguish healthy from malignant tissue with a spatial resolution of about 2 mm.

Philip Ball

Physics

Straight to the point

Science **305**, 1267–1269 (2004)

The properties of a light wave, such as its amplitude or wavelength, are easily probed. But the way in which its electric field evolves, or oscillates, has only now been measured.

E. Goulielmakis *et al.* used a laser to generate high-energy pulses of ultraviolet light that blasted electrons away from the atoms in a sample of neon gas. The light pulse lasted just 250 attoseconds (2.5×10^{-16} s) — not much longer than the time it takes the electron to orbit the proton in a Bohr-model hydrogen atom, and a tiny fraction of the oscillation period of visible light. As the electrons were ripped from the neon atoms, they acquired a certain momentum boost, which the authors measured. This momentum depends on what part of the light wave — peak, trough or somewhere in between — hits the electron.

Because an electric field is defined as the force exerted on a single point charge, this is the experimental realization of that theoretical definition. The authors note that their results fit perfectly with theoretical predictions, and suggest that this technique could be used to precisely control the quantum transitions of electrons within atoms and molecules.

Mark Peplow

Use of dung as a tool by burrowing owls

This bird distributes animal dung in and around its burrow to provide a bait for its prey.

Reports of tool usage by birds tend to be anecdotal as only a few individuals may be involved¹ and the behaviour observed can often be interpreted in other ways^{2,3}. Here we describe the widespread collection of mammalian dung by burrowing owls (*Athene cunicularia*) and show that they use this dung as a bait to attract dung beetles, a major item of prey. Our controlled investigation provides an unambiguous estimate of the importance of tool use in a wild animal⁴.

Burrowing owls place mammalian dung in and around their burrows⁵ (Fig. 1). If removed, it may be rapidly replaced, suggesting that it is more than an incidental accumulation of debris⁶. Because burrowing owls are one of the main predators of dung beetles^{5,7} and stand motionless for long periods by their burrows, we investigated whether they could be using mammalian dung as bait to 'fish' for dung beetles. The largest common species of dung beetle at our study site (*Phanaeus igneus*, 2 cm in length) is diurnal, as are the owls. This species makes up 65% of the beetles consumed by the owls ($n = 20$ owl pellets analysed). Alternatively, mammalian dung may mask the scent of the nest from predators⁶, given that the nests of burrowing owl are accessible to a wide range of terrestrial predators and that many nests are lost to them⁵.

To test this idea of olfactory camouflage, we created 50 nest burrows, spaced 50 m apart, and placed five quail eggs in each; every alternate burrow received cow dung. We recorded nest fate every two days for



Figure 1 Baiting and waiting: an owl stands at the entrance to its burrow, surrounded by the dung that it has positioned to entice beetles.

3.5 weeks, which is a typical incubation period. All but one nest was discovered and destroyed by predators. Survival analysis revealed no difference in time to destruction between nests with and without dung (Cox log-hazard ratio of 0.62 ± 0.31 , $z = 1.58$, $P > 0.10$). We conclude that dung does not effectively mask the scent of eggs. It may, however, be effective at masking the scent of chicks, which we did not test for ethical reasons.

To test the baiting hypothesis, we first removed all dung, regurgitated pellets and beetle parts from the burrow entrances of two owl populations. Half received 231 ± 16 g (dry mass) of cow dung, typical of the amount found at a burrow entrance; the remainder received no dung but were otherwise treated in the same way. After four days, we collected all prey remains and regurgitated pellets from the burrows and repeated the experiment, this time switching the treatment and control burrows. We found that when dung was present at the burrows, owls consumed ten times more dung beetles and six times more dung-beetle species than when dung was not present ($P < 0.001$; Fig. 2).

A few instances of tool use by wild birds are wonderfully detailed and widely accepted^{1,4}. More generally, however, the ecological and evolutionary significance of tool use is difficult to judge because evidence for how such behaviour benefits wild animals is scant⁴. Trials with captive animals have shown that they have surprising abilities⁸, but such results are difficult to extrapolate to natural settings. Baiting with dung and then waiting for dung beetles to come is akin to the use by herons of floating objects (such as bread, feathers and insects) as tools to attract fish⁹, although the foraging success with and without 'bait' has not been compared. Our experiment provides such a comparison, demonstrating that tool

use can substantially benefit a wild animal.

Douglas J. Levey*, **R. Scot Duncan†**,
Carrie F. Levins‡

*Department of Zoology, PO Box 118525,
University of Florida, Gainesville, Florida
32611-8525, USA

e-mail: dlevey@zoo.ufl.edu

†Division of Science and Mathematics,
Birmingham-Southern College, Box 549022,
Birmingham, Alabama 35254, USA

‡107 North El Centro Boulevard, Panama City
Beach, Florida 32413, USA

1. Hunt, G. R. *Nature* **379**, 249–251 (1996).
2. Heinrich, B. *Condor* **90**, 270–271 (1988).
3. Cristol, D. A., Switzer, P. V., Johnson, K. L. & Walke, L. S. *Auk* **113**, 296–298 (1997).
4. Tebbich, S., Taborsky, M., Fessl, B. & Dvorak, M. *Ecol. Lett.* **5**, 656–664 (2002).
5. Haug, E. A., Millsap, B. A. & Martell, M. S. (eds) *Burrowing Owl* (Acad. Natl Sci., Philadelphia, 1993).
6. Martin, D. J. *Condor* **75**, 446–456 (1973).
7. Woodruff, R. E. *The Scarab Beetles of Florida* (Florida Department of Agriculture and Consumer Services, Gainesville, 1973).
8. Weir, A. A. S., Chappell, J. & Kacelnik, A. *Science* **297**, 981 (2002).
9. Robinson, S. K. *Wilson Bull.* **106**, 567–569 (1994).

Competing financial interests: declared none.

Botany

A new self-pollination mechanism

Pollen grains from most flowering plants are transported by wind or animals and deposited on the receptive surface of the stigma of a different individual, but self-pollination is also common. We have discovered a new process for self-pollination in the laterally orientated flowers of a Chinese herb, in which a film of pollen is transported from the anther (pollen sacs) by an oily emulsion that slides sideways along the flower's style and into the individual's own stigma. This mode of

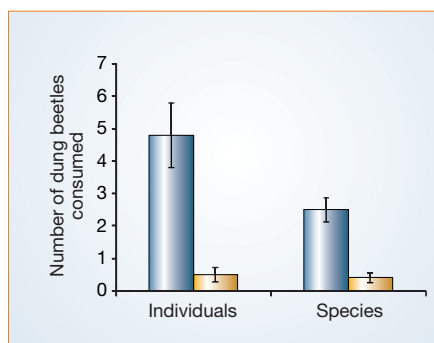


Figure 2 Number of dung beetles consumed by burrowing owls in the presence (blue) and absence (orange) of cow dung positioned at their burrows. Averages and standard errors are shown; $n = 10$ burrows per treatment in two populations (Gilchrist County, Florida). $P < 0.001$ for both the number of individuals and the number of species (in paired t -tests). Beetles were identified by their elytra, which owls commonly discard before consumption or regurgitate in pellets. Because we restricted our sampling of prey remains to the immediate vicinity of each burrow (to an area of about 0.75 m^2), the total number of beetles consumed during the trials will have been higher than indicated.

self-pollination is a new addition to the broad range of genetic and morphological mechanisms that have evolved in flowering plants (angiosperms)¹, and may be common in species growing in shady, windless and insect-poor habitats.

Caulokaempferia coenobialis (Hance) K. Larsen is a deciduous perennial herb of up to 50 cm in height. It is endemic in the Guangdong and Guangxi provinces of southern China, where it hangs on rock walls in humid monsoon forests^{2–4}. Plants flower from May to August and generally produce three buds that open consecutively. Flowers open between 6:00 and 6:30 in the morning, last for two days, and fade during the afternoon of the second day. Capsules mature and burst open, or dehisce, within 22 days. Seeds germinate and develop into seedlings that develop rhizomes before the plants die back between September and November, depending on the onset of the dry season. New plants develop from these rhizomes after the start of the next rainy season.

We studied the pollination biology of *C. coenobialis* in several large populations (2,000–10,000 flowering individuals) growing in the Dinghushan and Nankunshan nature reserves in Guangdong province in 2002, 2003 and 2004. *C. coenobialis* flowers have yellow corollas, with two short lateral lobes and one large central lobe of up to 3 cm long (Fig. 1a).

Although the precise phylogenetic affinities of *Caulokaempferia* within Zingiberaceae remain unresolved^{3,5}, the species has a flower structure typical of this family, with two elongated, lengthwise-dehiscing pollen sacs that enclose the style. The concave stigma lies almost exactly at the end of the pollen sacs (Fig. 1b). It resembles the stigmas of other Zingiberaceae, which are characterized by wet (secretory), papillose surfaces⁶. As in some other Zingiberales^{7,8}, the pollen grains in *C. coenobialis* are held together by pollen-connecting threads and have an abundance of oily pollenkit on their surface (as confirmed by staining with Sudan III and IV). Pollenkit is a mixture of hydrophobic components produced by the inner lining of the anther (tapetum). It consists of lipids and a species-specific admixture of carotenoids, proteins and carboxylated polysaccharides⁹. It is also abundant in other angiosperms, especially bee-pollinated varieties⁹.

The pollenkit in *C. coenobialis* is clear and rich in unsaturated lipids. It forms an oily film in which the pollen grains are suspended (Fig. 1c,d). Soon after anther dehiscence in the morning, a drop of pollen spills from each pollen sac. The two drops merge to form a film that quickly spreads across the style surface and slides towards the stigma (Fig. 1c,d). A fringe of hairs around the style may prevent the pollen from spilling off the style.

Spreading of the pollen film occurs

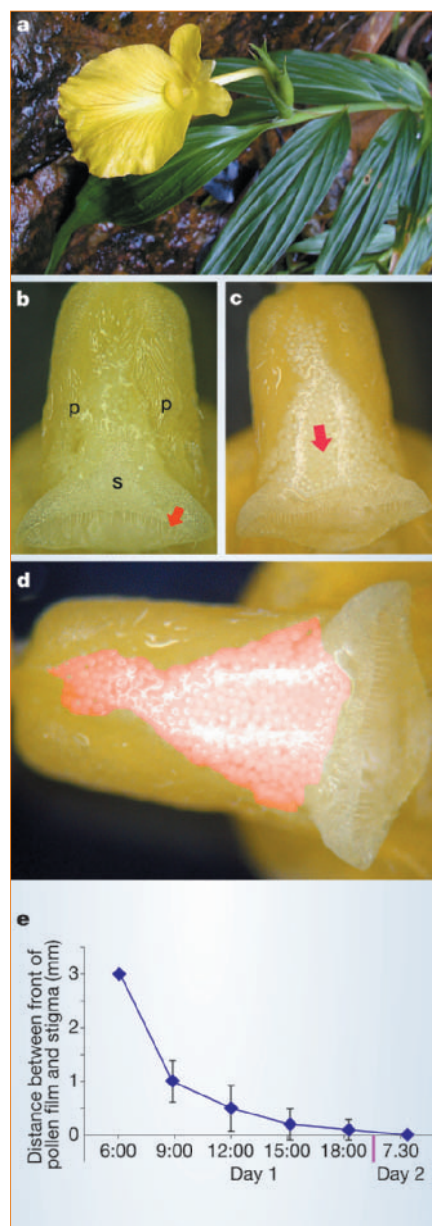


Figure 1 Pollen sliding in *Caulokaempferia coenobialis*. **a**, The flower in its natural position. **b**, The stigma (s) and the anther, consisting of two pollen sacs (p), which have pollen streams coming from them; arrow indicates stigma hairs. The stigma is about 3 mm in diameter. **c**, Pollen film sliding (arrowhead) towards the stigma. **d**, Pollen film stained red with Sudan III (computer reproduction of an actual experiment): the pollen film is on the style surface, about 0.2 mm from the stigma's margin; the pollen grains are visible in the oily film. **e**, Distance travelled over a period of about 25 hours by the pollen film as it flows from the pollen sac towards the stigma (distances are measured between the pollen-film front and the stigma; mean values $\pm$ s.d.; $n = 20$).

rapidly (Fig. 1e), the rate being comparable for plants in the laboratory, at 25–65% relative humidity, and plants in their nearly saturated natural habitat on outcrops in forests. The film arrives at the stigma margin, where there is a row of fine hairs, and passes between them at about 15:00, at which point self-pollination begins and continues until the next morning.

The rate of fruit set in five populations fol-

lowing controlled hand self-pollination ($95.93 \pm 4.77\%$, $n = 49$) and cross-pollination ($97.96 \pm 4.97\%$, $n = 49$), in which a drop of pollen was placed directly on the stigma, does not differ significantly ($t = 0.6622$, $P > 0.05$), indicating that *C. coenobialis* is capable of selfing without detrimental effects on seed set. There was also no significant difference ($t = 0.1717$, $P > 0.05$) between the rate of fruit set in control flowers and in flowers bagged before opening ($75.93 \pm 22.37\%$, $n = 50$, compared with $77.89 \pm 12.63\%$, $n = 95$). The species is incapable of asexual seed set, as shown by the absence of seed set in flowers that had had their stamens or stigmas cut off before they could self-pollinate ($n = 50$ for each experiment).

To our knowledge, this is the first report of pollination in angiosperms by pollen that is conveyed in a mobile secreted medium. The lateral flow of the film of pollen along the style seems to be due only to the spreading properties of the oily emulsion and not to gravity. Analysis of the style and stigma under a scanning electron microscope reveals a smooth style and stigma surface, and no channels or grooves. During more than 20 days of observation (about 200 hours), we never witnessed any insects visiting the flowers that would have touched the anthers and stigmas, even though each flower can offer up to two microlitres of nectar containing 18% sugar. Selfing by virtue of pollen sliding to the stigma may have evolved as a strategy to cope with a scarcity of pollinators in the extremely shady and humid habitats of *C. coenobialis*.

Yingqiang Wang*†, Dianxiang Zhang*, Susanne S. Renner‡, Zhongyi Chen*

*South China Botanical Garden, The Chinese Academy of Sciences, Guangzhou 510650, China
e-mail: dx-zhang@scib.ac.cn

†Zhongkai Agrotechnical College, Guangzhou 510225, China

‡Department of Biology, Ludwig Maximilians University, 80638 Munich, Germany

- Barrett, S. C. H. *Phil. Trans. R. Soc. Lond. B* **358**, 991–1004 (2003).
- Larsen, K. *Botaniska Tidskrift* **60**, 165–179 (1964).
- Larsen, K. *Nord. J. Bot.* **22**, 409–417 (2002).
- Larsen, K. & Smith, R. M. *Notes R. Bot. Gard. Edinb.* **31**, 287–295 (1972).
- Williams, K. J., Kress, W. J. & Manos, P. S. *Am. J. Bot.* **91**, 100–114 (2004).
- Heslop-Harrison, J. & Shivanna, K. R. *Ann. Bot.* **41**, 1233–1258 (1977).
- Kronstedt, E. & Byström, P. A. *Nord. J. Bot.* **1**, 523–529 (1981).
- Rose, M. J. & Barthlott, W. *Plant Syst. Evol.* **195**, 61–65 (1995).
- Dobson, H. E. M. *Am. J. Bot.* **75**, 170–198 (1988).

Competing financial interests: declared none.

brieF communications arising online

www.nature.com/bca

Superconductors: Time-reversal symmetry breaking?

S. V. Borisenko, A. A. Kordyuk, A. Koitzsch, M. Knupfer, J. Fink, H. Berger & C. T. Lin (doi:10.1038/nature02931)

Reply: J. C. Campuzano, A. Kaminski, S. Rosenkranz & H. M. Fretwell (doi:10.1038/nature02932)

Superconductors

Time-reversal symmetry breaking?

Arising from: Kaminski, A. *et al.* *Nature* **416**, 610–613 (2002)

One of the mysteries of modern condensed-matter physics is the nature of the pseudogap state of the superconducting cuprates. Kaminski *et al.*¹ claim to have observed signatures of time-reversal symmetry breaking in the pseudogap regime in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (Bi2212). Here we argue that the observed circular dichroism is due to the 5×1 superstructure replica of the electronic bands and therefore cannot be considered as evidence for spontaneous time-reversal symmetry breaking in cuprates.

The main conclusions of Kaminski *et al.* are based on the temperature-dependent circular dichroism observed in a ‘mirror’ plane of the underdoped Bi2212 thin films. However, Bi2212 samples possess a 5×1 superstructure that breaks reflection symmetry in these planes, as demonstrated by electron diffraction and angle-resolved photoemission (ARPES) experiments (Fig. 1a,b). The 5×1 superstructure is suppressed by doping pristine Bi2212 with lead. We performed ARPES experiments similar to those reported by Kaminski *et al.*¹ on both systems. At room temperature (Fig. 1c), the influence of the superstructure is already obvious: for

pristine Bi2212, the dichroic signal is non-zero in the $\Gamma-(\pi,0)$ plane.

This result can readily be explained. The superstructure results in diffraction replicas of the electronic structure seen in the momentum-distribution map (Fig. 1b) as green and blue dashed curves. Because of the pronounced inequivalence of the matrix elements in the first and second Brillouin zones, the spectral weight of these replicas is always different near the $(\pi,0)$ point. Recording the dichroism as a function of momentum $\mathbf{k}$ along the white double-headed arrow in Fig. 1b, one effectively measures the superposition of the three signals originating from the main band and two non-equivalent diffraction replicas. A systematic investigation of the 5×1 superstructure-free Pb-Bi2212 samples that have a large pseudogap (Fig. 1d) reveals that the dichroism in the mirror plane remains zero within the experimental error bars, independent of temperature (Fig. 1c) and doping².

The finite superstructure-induced room-temperature dichroism in the mirror plane must also be present in the thin films, which do exhibit a superstructure signal. The energy-distribution curves (EDCs), which are

identical to an accuracy of 0.06% (see Fig. 2e of ref. 1), therefore indicate that they cannot be taken at the $(\pi,0)$ point and that the zero momentum in Fig. 3g of ref. 1 probably does not correspond to the mirror plane. The large uncertainty in locating the actual position in the $\mathbf{k}$ -space (not specified by Kaminski *et al.*) is also evident from data shown in their Fig. 3d,h. Presented EDCs for the overdoped sample are claimed to be $\mathbf{k}_f$ EDCs. However, it is known that, at finite temperature at $\mathbf{k}_f$, the spectral function has a peak at the chemical potential and so multiplication by the Fermi function would result in the leading-edge midpoint being located at negative binding energies, which is not the case.

Provided that the zero momentum in Fig. 3g of ref. 1 does not correspond to the $(\pi,0)$ point, the temperature dependence of the dichroism is not surprising. Away from the mirror plane, the dichroism corresponding to the main band is temperature dependent, as can be seen by comparing the slopes of the dichroism in their Fig. 3c (note that the lines shown in Fig. 3c,g of ref. 1 are not always linear fits to the 11 data points, as is evident from the data collected at 150 K) and in Fig. 3d of ref. 2.

The absence of full-range curves² in Fig. 3c, g of ref. 1 does not allow the exact location in momentum space from which the presented data are taken to be determined; neither can we determine whether this is

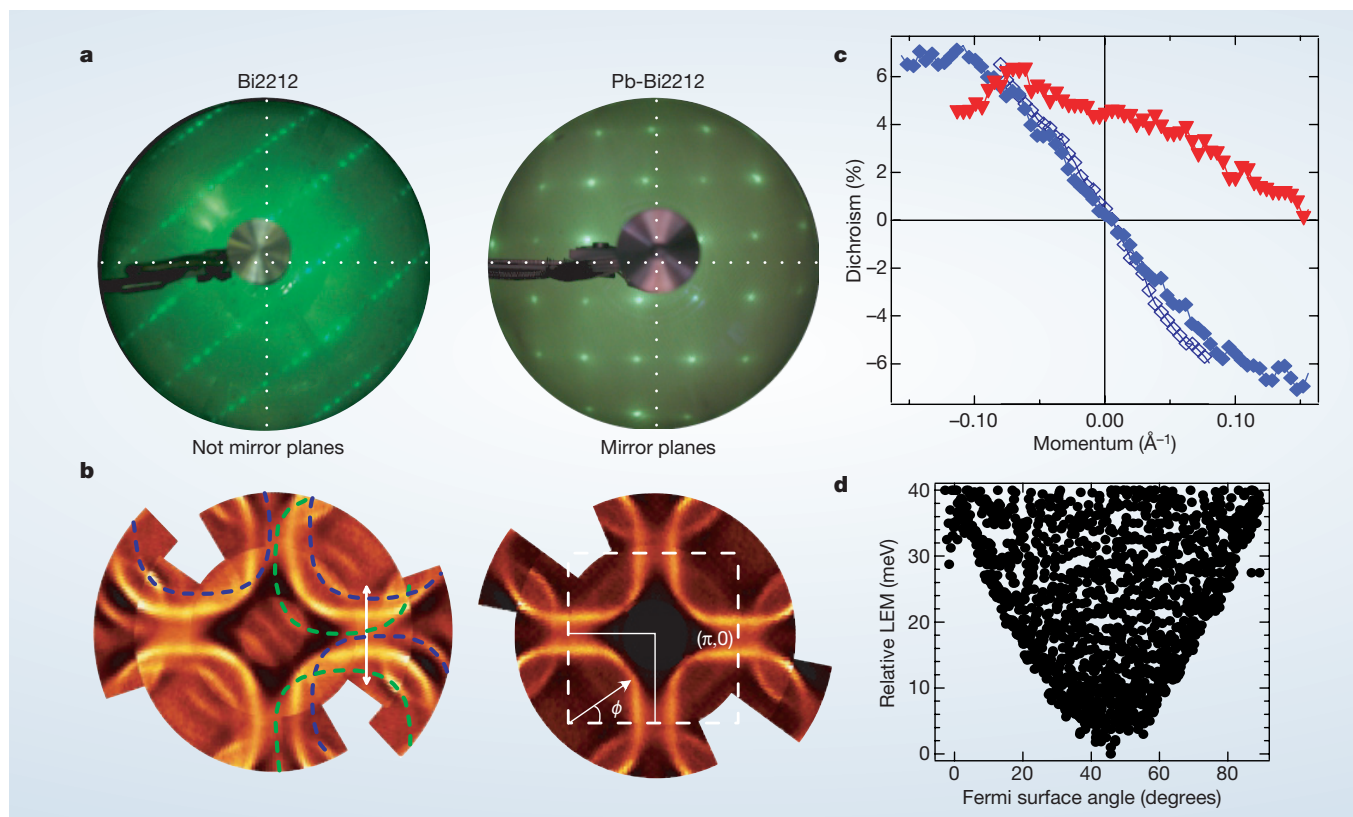


Figure 1 Features of superstructure in superconducting cuprates. **a, b**, Electron diffraction (**a**) and angle-resolved photoemission (**b**), showing angular distributions of the electrons in pristine (left) and Pb-doped (right) Bi2212. White dotted lines, crystallographic planes; green and blue dashed lines, diffraction replicas; white dashed line, first Brillouin zone. **c**, Dichroism near $(\pi,0)$ in Bi2212 (red triangles) and Pb-Bi2212 (blue diamonds; filled, 300 K; open, 100 K). **d**, Anisotropic pseudogap in Pb-Bi2212 measured at 120 K as binding energy of the leading-edge midpoints (LEM) for all spectra within the quadrant of the Brillouin zone as a function of Fermi surface angle ϕ (white arrow in **b**, right).

always from the same place. In support of such uncertainty, there is a considerable (about 14%) variation in the slope of the 250 K 'line' in two similarly underdoped samples (see Figs 3g, 4a,b of ref. 1) that cannot be explained by the small difference in doping levels, because comparison with a much more strongly doped sample (Fig. 3c of ref. 1) gives a similar change in slope.

The results presented in Fig. 4a,b of ref. 1 also fit the 'superstructure' scenario: in this, going from the $(\pi,0)$ to the $(0,\pi)$ point (our Fig. 1), the stronger (blue) diffraction replica is now on the other side of the $(0,\pi)$ point and therefore the temperature-induced changes have the opposite sign.

The data of Kaminski *et al.* on the overdoped sample (their Fig. 3c) can be explained by the substantially weaker influence of the diffraction replica in the immediate vicinity of the $(\pi,0)$ point because of the larger size of the Fermi surface (we consider only the bonding sheet, as the antibonding one is strongly suppressed at the photon energies used in ref. 1; Fig. 1b). In addition, the superlattice signal seems to be more sensitive to the temperature in underdoped samples³ and to vary from sample to sample: Kaminski *et al.* report it to be about 3%, whereas it is about 10% in ref. 4.

Sergey V. Borisenko*,
Alexander A. Kordyuk*†,
Andreas Koitzsch*, Martin Knupfer*,
Jörg Fink*, Helmuth Berger‡,
Chengtian T. Lin§

*Institute for Solid State Research, IFW–Dresden,
PO Box 270016, 01171, Dresden, Germany
e-mail: s.borisenko@ifw-dresden.de

†Institute of Metal Physics, 03142 Kyiv, Ukraine

‡Institute of Physics of Complex Matter, EPFL,
1015 Lausanne, Switzerland

§Max-Planck Institute for Solid State Research,
70569 Stuttgart, Germany
doi:10.1038/nature02931

1. Kaminski, A. *et al.* *Nature* **416**, 610–613 (2002).
2. Borisenko, S. V. *et al.* *Phys. Rev. Lett.* **92**, 207001 (2004).
3. Armitage, N. P. & Hu, J. *Phil. Mag. Lett.* **84**, 105–107 (2004).
4. Kaminski, A. *et al.* Preprint at <http://arXiv.org/cond-mat/0306140> (2003).

Kaminski *et al.* reply — There are two components of the circular dichroism (CD) signal in angle-resolved photoemission (ARPES) measurements. One is always present in crystals, regardless of any time-reversal symmetry considerations. This component, which we refer to as 'geometric', is antisymmetric about any symmetry plane of the crystal, and is therefore zero at that plane. But in underdoped samples of the high-temperature superconductor Bi2212, we find another component, which is non-zero at the symmetry plane below the pseudogap temperature. We attribute that component to time-reversal symmetry breaking. The objections of Borisenko *et al.*¹ comprise three main points: the circular

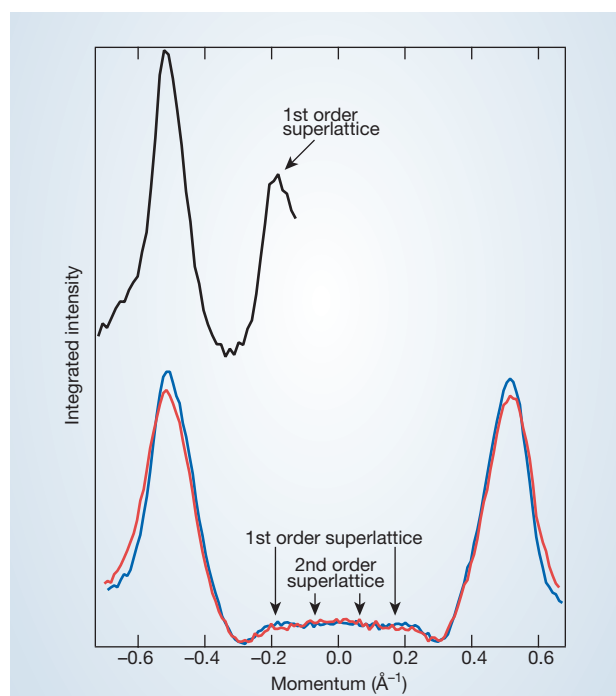


Figure 1 Comparison of the superstructure signal and its effect on the circular dichroism in single-crystal samples and epitaxially grown thin films. The momentum distribution data along the $(0,0)-(\pi,\pi)$ direction obtained from a single crystal sample (top black curve) displays a superstructure signal of about 50%. Similar data obtained from epitaxially grown thin films show only a 3% superstructure signal (lower curves: red, 200 K and blue, 50 K).

dichroism that we observe at the mirror plane² is due to the superstructure of the Bi–O layer; our momentum accuracy is not as we stated; and the absence of dichroism in overdoped samples is due to a weaker influence of the superstructure because of an increased Fermi surface volume compared with underdoped samples.

It is obvious that modulation of the BiO plane breaks reflection symmetry about the Cu–O bond direction in BSCO crystals (that is, those containing Bi, Sr, Ca, Cu, O), but our experiment relies on the fact that the CuO₂ states do not couple to the BiO states, thereby preserving this symmetry³. We explicitly test this in Fig. 3f of ref. 2, where we show that there is no dichroism at high temperature at the mirror plane in a sample, but that dichroism does occur below the pseudogap temperature.

The first point of Borisenko *et al.* could only be valid if two conditions were met simultaneously: the superstructure signal would have to be large (about 40%), and either the geometric component of the CD signal or the superstructure signal would have to display significant temperature dependence. The first condition does not hold, as shown by our measurements on epitaxially grown thin films that have a very weak superstructure signal^{2,4} (of the order of 3%; Fig. 1). We agree that this signal in single-crystal samples was about 30–50% of the main signal, and we showed earlier^{3,5,6} that this was entirely due to the diffraction of the outgoing photoelectron. If the dichroism originates from

the superstructure signal, then the 4% dichroism resulting from the 40% superstructure signal in the crystal of Borisenko *et al.*¹ would indicate that our 3% superstructure signal would produce a dichroism of only 0.3% — that is, we would never observe a signal.

The second of the two conditions also fails, as there is strong evidence that the geometric component of the CD does not depend on temperature. The slope of the CD at the mirror plane is due entirely to the anti-symmetric component and does not change with temperature (see Fig. 3c,g of ref. 2 and Fig. 4b of ref. 7). More significantly, if the superstructure signal contribution is large, then the dichroism signal would be a nonlinear function of momenta in the vicinity of the $(\pi,0)$

point. Such nonlinear behaviour would be significant, as demonstrated by a simulation of the CD that assumes a 40% level of superstructure signal contamination (results not shown), but this is not seen in our data, which instead resemble the simulation result for a 4% superstructure signal (results not shown). Also, as the superstructure signal does not depend on temperature (Fig. 1), the temperature-dependent CD cannot originate from the superstructure signal.

The second point raised by Borisenko *et al.* is based on a misconception. The midpoint of the leading edge lies at positive binding energies only for a very sharply peaked spectral function. For a broad function, it lies close to the chemical potential. It is clear that the areas beneath the curves in our Fig. 3d (ref. 2) are independent of temperature, which holds only at the Fermi momentum^{8,9}. Both our samples and detector were aligned with precision².

The third point assumes that the Fermi momentum along the $(\pi,0)-(\pi,\pi)$ direction changes significantly with doping. Such a change would have to be larger than at least half of the momentum range measured in our experiment (our momentum range was 0.1 Å^{-1}) to produce the alleged effect. This is because in underdoped samples the superstructure would have to be well within our momentum range, and in overdoped samples it would have to lie outside. Other data of Borisenko *et al.*¹⁰ provide the best evidence that such large shifts are not observed experimentally. For example, Fig. 1 of ref. 10 shows that the Fermi momentum for both

overdoped and underdoped samples is similar and equal to $0.15 \text{ \AA}^{-1}$.

The explanation for our results proposed by Borisenko *et al.* fails to account for both the temperature and doping dependence of the dichroism signal at the mirror plane. Their own data from underdoped samples (Fig. 3c in ref. 7) shows temperature-dependent dichroism at the symmetry plane of about 1%. As those data were obtained

from superstructure-signal-free samples, the superstructure cannot be responsible for the observed signal.

**Juan C. Campuzano, Adam Kaminski,
Stephan Rosenkranz, Helen M. Fretwell**
*Argonne National Laboratory, Argonne, Illinois
60439, USA*

*Department of Physics, University of Illinois at
Chicago, Chicago, Illinois 60607, USA*
e-mail: campuzano5@comcast.net

doi:10.1038/nature02932

1. Borisenko, S. V. *et al.* *Nature* doi:10.1038/nature02931 (2004).
2. Kaminski, A. *et al.* *Nature* **416**, 610–613 (2002).
3. Norman, M. R. *et al.* *Phys. Rev. B* **52**, 615–622 (1995).
4. Kaminski, A. *et al.* Preprint at <http://arXiv.org/cond-mat/0306140> (2003).
5. Ding, H. *et al.* *Phys. Rev. Lett.* **76**, 1533–1536 (1996).
6. Fretwell, H. M. *Phys. Rev. Lett.* **84**, 4449–4452 (2000).
7. Borisenko, S. V. *et al.* *Phys. Rev. Lett.* **92**, 207001 (2004).
8. Sato, T. *et al.* *Phys. Rev. B* **64**, 054502 (2001).
9. Mesot, J. *et al.* *Phys. Rev. B* **63**, 224516 (2001).
10. Kim, T. K. *et al.* *Phys. Rev. Lett.* **91**, 167002 (2003).

Active foundering of a continental arc root beneath the southern Sierra Nevada in California

George Zandt¹, Hersh Gilbert¹, Thomas J. Owens², Mihai Ducea¹, Jason Saleeby³ & Craig H. Jones⁴

¹Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA

²Department of Geological Sciences, University of South Carolina, Chapel Hill, North Carolina 29208, USA

³Department of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

⁴Department of Geological Sciences, University of Colorado, Boulder, Colorado 80309, USA

Seismic data provide images of crust–mantle interactions during ongoing removal of the dense batholithic root beneath the southern Sierra Nevada mountains in California. The removal appears to have initiated between 10 and 3 Myr ago with a Rayleigh–Taylor-type instability, but with a pronounced asymmetric flow into a mantle downwelling (drip) beneath the adjacent Great Valley. A nearly horizontal shear zone accommodated the detachment of the ultramafic root from its granitoid batholith. With continuing flow into the mantle drip, viscous drag at the base of the remaining ~35-km-thick crust has thickened the crust by ~7 km in a narrow welt beneath the western flank of the range. Adjacent to the welt and at the top of the drip, a V-shaped cone of crust is being dragged down tens of kilometres into the core of the mantle drip, causing the disappearance of the Moho in the seismic images. Viscous coupling between the crust and mantle is therefore apparently driving present-day surface subsidence.

Giant magmatic bodies, or batholiths, that constitute nearly the entire crust in belts thousands of kilometres long are integral features of major mountain systems associated with continental arcs. There is a growing recognition that the growth of ~30-km-thick granitoid batholiths must involve a two-step magmatic differentiation process that generates an equal or even greater thickness of ultramafic residual roots beneath the granitic bodies (see, for example, ref. 1). Removal of such dense ultramafic roots is often cited as a mechanism for differentiation in Cordilleran arcs (see, for example, refs 2, 3) to produce the overall intermediate composition of the continents^{4,5}, but the actual timing and mechanism of removal, and whether all such roots eventually founder, remains uncertain. In southern California, a compelling case has been made for a late Cenozoic (<10 Myr ago) foundering event beneath the southern Sierra Nevada batholith^{6–10}. We present seismic results that provide direct evidence connecting crustal structures associated with the removal event to the mantle downwelling imaged in previous tomography studies^{11–13} (Supplementary Information). Analyses of receiver functions reveal seismic anisotropy at the base of the crust that may be related to shearing along a detachment zone that became a new crust–mantle boundary. Viscous drag at the base of the remaining ~35-km-thick crust has thickened it by ~7 km in a narrow crustal welt, and at the top of the drip, a V-shaped cone of crust is being dragged down tens of kilometres into the core of the mantle drip.

Crustal welts and missing Moho

A 20-station, PASSCAL experiment (in 1997) and five permanent stations in the TriNet network in this area recorded the teleseismic data used in this study (Fig. 1). The 25 stations encompass the southern Sierra Nevada region, where xenoliths from Miocene, Pliocene and Quaternary volcanic fields document the presence of an ultramafic root in the Miocene (~10 Myr ago) and its absence in the Pliocene (~3 Myr ago)^{6–8}. Residual garnet pyroxenite xenoliths found in mid-Miocene volcanic rocks from the Sierra Nevada⁶ are extremely dense rocks (3,500–3,600 kg m⁻³) compared with typical mantle peridotites (~3,300 kg m⁻³) owing to their garnet- and Fe-rich nature⁴. This heavy residue is prone to separation from its

overlying low-density granitoid batholith and eventual convective foundering into the mantle. A pulse of small-volume, high-potassium volcanism that erupted within a circular area about 200 km in diameter centred just south of Long Valley (Fig. 1a) apparently indicates both the locality and timing of the main removal event in the southern Sierra Nevada^{14,15}. This might represent only part of the area where the root was removed, as 3–4-Myr-ago volcanism extends much farther north and south¹⁶.

Receiver function seismology uses seismic waves produced by P- to S-wave mode conversions generated below seismic stations to image intracrustal and upper-mantle interfaces. Like reflection seismology, receiver functions are seismic traces that can be migrated to depth and geographically stacked to make pseudo-cross-sections of the subsurface in which the depths of continuous horizons, such as the Moho, can be identified and mapped. We stacked receiver functions calculated from teleseismic recordings in common conversion area bins (see, for example, refs 17–19). A crustal thickness map based on contouring the Moho P-to-s-converted phase (Ps) reveals a complex pattern of thickness variations in the Sierra Nevada and surrounding regions (Fig. 1b). Beneath the northeastern corner of the study area, which includes the Big Pine volcanic field, the crustal thickness is about 36 km, 7–8 km thicker than the crust in Owens Valley farther south. The crust thickens westward, forming a crustal welt with a maximum thickness of 42 km beneath the Kings volcanic field. The Moho changes to a very different character south of the area of the high-potassium volcanism, where the crust in the Sierra Nevada is almost everywhere less than 36–37 km thick. Under the southern Owens Valley and adjacent ranges to the east, the Moho is relatively flat at ~30 km.

The pattern of crustal thickness variations determined from receiver functions agrees well with the results from seismic refraction studies²⁰. In particular, both indicate thicker crust (~40 km) beneath the Sierra Nevada north of about 36.5° latitude. The main difference between the two models that these methods produce is that the refraction model shows a bulge of thicker crust (~35 km) beneath the Owens Valley extending from south of Big Pine to about

36° latitude that is absent in the receiver function Moho-depth model (~30 km). Our crustal thicknesses are also in good agreement with an independent receiver function study done with the permanent stations in the Caltech network that overlap our study area²¹.

Three northeast–southwest receiver function cross-sections are shown in Fig. 2, with interpretations overlaid to highlight some interesting details of the crustal structure. The Moho is a large-amplitude, positive-polarity (red) arrival at depths between 30 and 42 km. In the southern cross-section (A–A') the Moho is a smooth interface, deepening gradually to the west, where the amplitude diminishes abruptly beneath the Great Valley. In the middle cross-section (B–B') the Moho arrival disappears farther east beneath the western edge of the range. In the northern cross-section (C–C') the Moho arrival is more complex, exhibiting en echelon offsets before disappearing even farther east beneath the western foothills. An absence of a clear wide-angle Moho reflection (PmP) from this portion of the crust was noted in the active source experiments, but it was attributed to noisy recording conditions in the Great Valley^{13,20}. In contrast, all the broadband stations were located on crystalline bedrock, and examination of data from the six stations sampling the region of absent Moho, some of which extend well into the western flank of the range, show that the disappearance is not due to interference from basin-generated noise.

The missing Moho in the western foothills (Fig. 1b) has two possible explanations. One scenario is that the shear velocity contrast at the Moho is significantly reduced by a localized zone of low S-wave velocity in the uppermost mantle, perhaps due to serpentinization of the mantle like that found in active forearc regions²². This is a plausible explanation, given that this region was within the forearc of a subduction zone until ~20 Myr ago. An alternative explanation is that small-scale topography on the Moho is scattering away the coherent Ps conversion, as well as the active

source PmP reflections. Later, we present synthetic modelling and additional independent evidence that supports this latter explanation.

Anisotropy in the lower crust

Another prominent feature in the cross-sections is a series of three en echelon negative-polarity arrivals (blue) observed in the middle to lower crust (at depths of ~20–30 km) throughout the study region (Fig. 2). This feature corresponds to the mid-crustal negative-polarity arrival (MCN) imaged in ref. 23 based on data from several small arrays deployed in an earlier study (MK, HM and DP00 in Fig. 1b). In the southern part of the study area, the top of this feature under Owens Valley is at a depth of ~20 km. Moving westward into the range, there appears to be a ~8–10 km en echelon step downward of the negative-polarity phase beneath the Sierra Nevada (A–A'). This 25-km-deep arrival underlies the entire high Sierra within the study area. Farther north (B–B'), the en echelon offset occurs beneath the eastern edge of the range. In the northern section where the crust thickens to a maximum of ~42 km, the thickening appears to occur in an en echelon fashion accompanied by another negative-polarity phase at ~35 km depth (C–C').

Analysis of directional variations in the negative-polarity arrival on the radial receiver function and its associated arrival on the tangential receiver function suggests the presence of an anisotropic layer at the base of the crust composed of an east-dipping fabric. The seismic anisotropy of this fabric is characterized by a slow direction perpendicular to an east or southeast-dipping plane, and equally fast directions in all orientations within that plane. One example of the azimuthal variations that results from the presence of anisotropy is displayed in the radial and tangential receiver functions for station JUN shown in Supplementary Information. We find that a planar fabric striking north–south to northwest–southeast and dipping at some amount greater than ~45° downward to the east or southeast

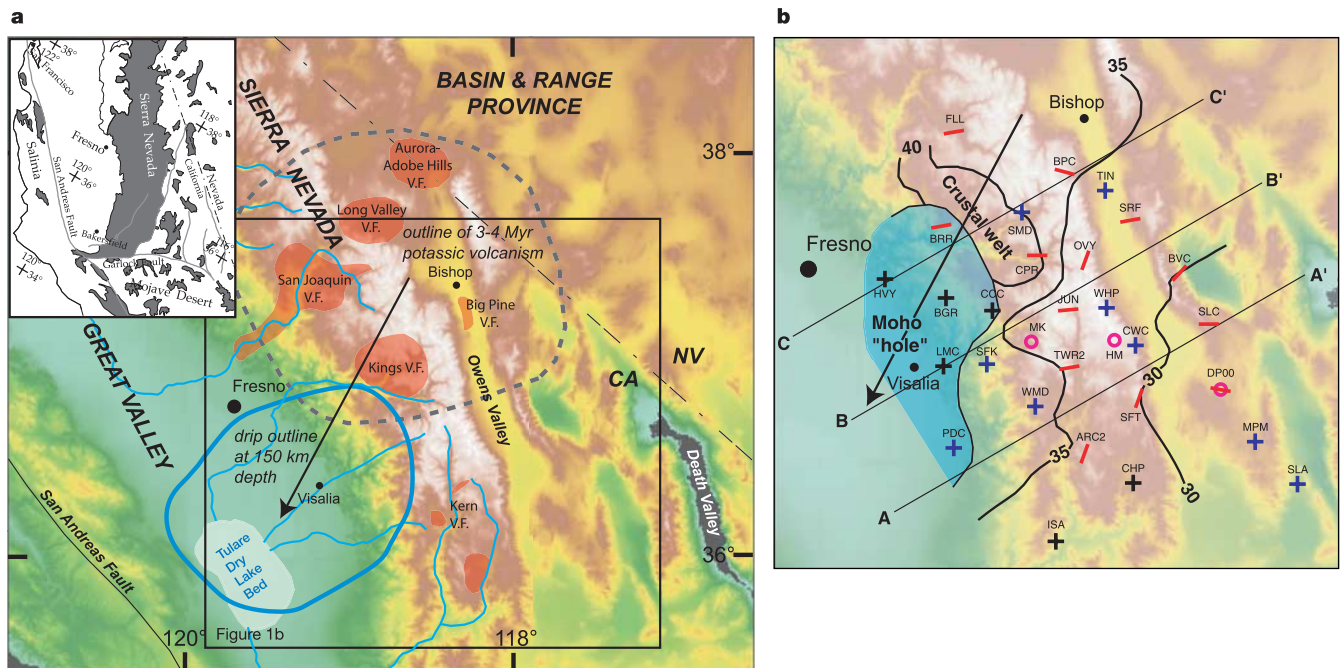


Figure 1 Study area and major results. **a**, Map of study area illustrating topography and drainage patterns, Cenozoic volcanic fields^{15,16,46}, outline of 3–4-Myr-old potassic volcanism¹⁴ and outline of the mantle drip at 150-km depth¹³. Inset shows distribution of granitoid batholithic rocks in the Sierra Nevada, Mojave Desert and the Salinian terrane. **b**, Locations of broadband seismic stations used in this study are marked with symbols and station codes, indicating the presence and direction (short red dash) of lower-crustal anisotropy, absence of anisotropy (black 'plus' sign), or presence of some but differently

oriented anisotropy (blue 'plus sign'). Orientation of the dash is along the slow-wavespeed direction and perpendicular to the fabric orientation. Small open circles locate small arrays used in an earlier receiver function study²³. Thick black lines are contours of Moho depths (in km) estimated from stacked receiver functions. Thin black line with blue transparency outlines region with no Moho arrival in the stacked receiver functions that is interpreted as a Moho hole (see text for further explanation). Three cross-sections (A–A', B–B', C–C') are shown in Fig. 2.

provides the best fit to our data. Azimuthally varying patterns similar to those shown in Supplementary Information are found at most stations located in the Sierra Nevada and adjacent ranges across Owens Valley. Stations showing either different patterns or little evidence for crustal anisotropy are largely located within or near the edge of the Moho 'hole', near the eastern front of the Sierra Nevada, or south of latitude 36° N where the root was removed earlier (Fig. 1b).

Resolution tests

The common-conversion-point stacking technique employed here assumes that the P-to-S-converted phases are produced by laterally continuous horizontal structures, and does not account for focusing

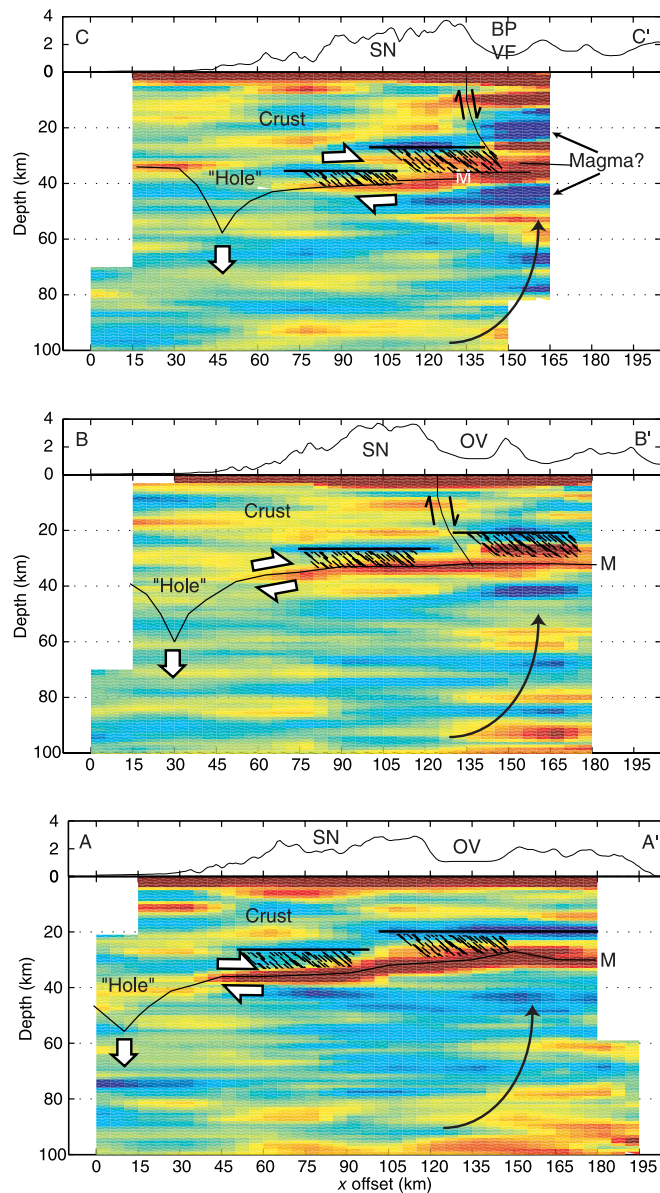


Figure 2 Cross-sections through the three-dimensional model of stacked receiver functions migrated into the depth domain and associated topography. Top panel, C–C'; middle, B–B'; bottom, A–A'; cross-sections are shown in Fig. 1b. Red areas represent large positive amplitudes (increase in wavespeed with depth), and blue areas represent negative amplitudes (decrease in wavespeed with depth). Hachure beneath prominent blue layers in crust represents east-dipping planar fabric producing seismic anisotropy. The V-shaped Moho holes are discussed in the text. SN, Sierra Nevada; BPVF, Big Pine volcanic field; OV, Owens Valley; M, Mono.

or diffraction effects resulting from dipping or laterally varying interfaces. Images presented here display a dip on the Moho as well as negatively polarized arrivals that appear as a series of en echelon steps. We test the veracity of these images by determining whether or not we are able to resolve features with dips or steps with coarsely spaced stations. These tests are performed by computing synthetic receiver functions using a finite-difference algorithm (O. Boyd, personal communication) for various structures, and then processing the synthetics in the same way as the data. The station spacing and incidence angles of events in the synthetic tests are chosen to reflect those of the actual data. First, the presence of a structural cusp on the Moho was modelled to see how such a structure would appear in receiver function stacks (Fig. 3). The size of the cusp (50 km wide by 25 km deep) was made to be similar to the volume of crustal material entrained into the mantle by the 'drip' as it appears in the cross-sections and was based loosely on published numerical models. This is the feature responsible for creating the observed 'hole' in the Moho. In the cusp model, significantly diminished amplitude arrivals are observed centred over the cusp, demonstrating that a cusp-shaped feature can produce the observed Moho hole. No clear arrivals off the sides of the cusp were seen. The dipping edges of the cusp are too steep to be imaged here, but energy does arrive at the base of the cusp. The tests demonstrate that a V-shaped notch of crust ~ 50 km wide by ~ 25 km deep, cut into the mantle, is sufficient to make the Moho disappear in the receiver function stack. This is the basis for the V-shaped Moho holes interpreted in the cross-sections of Fig. 2. The western edge of the hole is poorly constrained owing to the absence of stations in the Great Valley. Second, we tested a series of models with a dipping Moho overlaid by an en echelon low-velocity zone with vertical offsets of varying magnitude (see Supplementary Information). We found that with our station spacing en echelon steps offset by at least 5 km can be recovered, while smaller steps can appear as a continuously dipping layer. Additionally, tests show that a continuously dipping low-velocity zone is not aliased into a series of steps.

Alignment of crustal welt, Moho hole and mantle drip

The structural features and fabric of the crust that we have imaged do not correlate in any simple way with Sierran topography or geology. However, a revealing clue to the significance of these patterns is the alignment of major features along a southwest trend from near Bishop to Visalia (Fig. 1). Along this trend are the centre of the region of potassic volcanism, the crustal welt beneath the Kings volcanic field, the Moho hole beneath the adjacent foothills, the centre of the surface projection of the mantle drip and the Tulare dry lakebed. The final clue that provides a

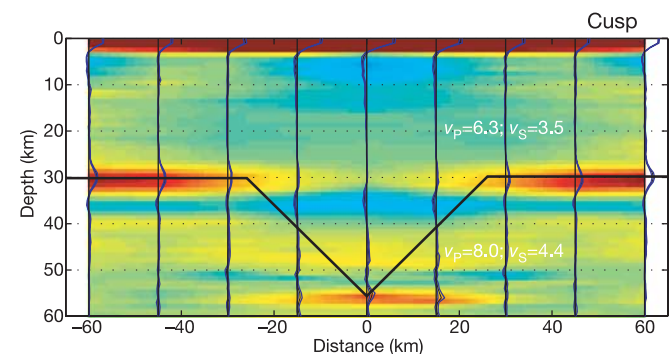


Figure 3 Synthetic stacked receiver function cross-section generated with a finite-difference wave-propagation technique for the seismic model shown beneath the image. The synthetic data were processed with station and bin spacings similar to those used with the southern Sierra Nevada data. v_P , P-wave velocity in km s^{-1} ; v_S , S-wave velocity in km s^{-1} .

connection among all these features is the recognition of a subsiding sub-basin in the Great Valley centred above the mantle anomaly²⁴. The locus of subsidence can be observed in the drainage patterns in the southern Sierra Nevada, where rivers discharging into the valley north of Fresno flow northward, while those within the circle of influence of the drip flow into the Tulare internal drainage basin (Fig. 1a). Ongoing work reconstructing sedimentation patterns suggests that the local sub-basin began forming at ~3–4 Myr ago and that it has migrated southwest to its current position²⁴. These observations can be connected to an active southwest-migrating convective instability. The alignment of features in the Sierra Nevada and the absence of similar features to the west suggest a strong asymmetry to the process. The localized crustal welt beneath the Kings volcanic field may reflect viscous drag at the base of the crust in the wake of southwest mantle motion (see, for example, refs 25–28). Numerical models of analogous convective instabilities^{29,30} show that at the point of downward detachment of the lithosphere, the lower crust will be entrained several tens of kilometres or more into the downward flow, providing an explanation for the V-shaped

Moho hole. The surface expression of the crustal foundering would be an elliptical zone of subsidence produced by viscous coupling between the downwelling drip and the overlying crust³¹.

We believe that the anisotropic fabrics observed near the crust–mantle boundary are associated with a shear zone between the crust and the foundering root. Well-developed shear zones generally have fabrics that are parallel to the shear zone, whereas fabrics or layering at an oblique angle to the shear plane are usually associated with the initial stages of shearing under brittle-ductile conditions³². The observed anisotropy could be due either to alignment of slow-axis minerals, such as mica, or to fluid- or melt-filled cracks in planes perpendicular to the axis direction. In both cases the planes containing the minerals or cracks dip east to northeast. The distinction between these two origins for the anisotropy is important, because they predict the opposite sense of shear with the same alignment (see Supplementary Information). The preferred interpretation is that the dipping fabric is related to fluid- or melt-filled fractures or veins formed during the initial stage of shearing. Cracks, fractures, or veins formed during shearing will

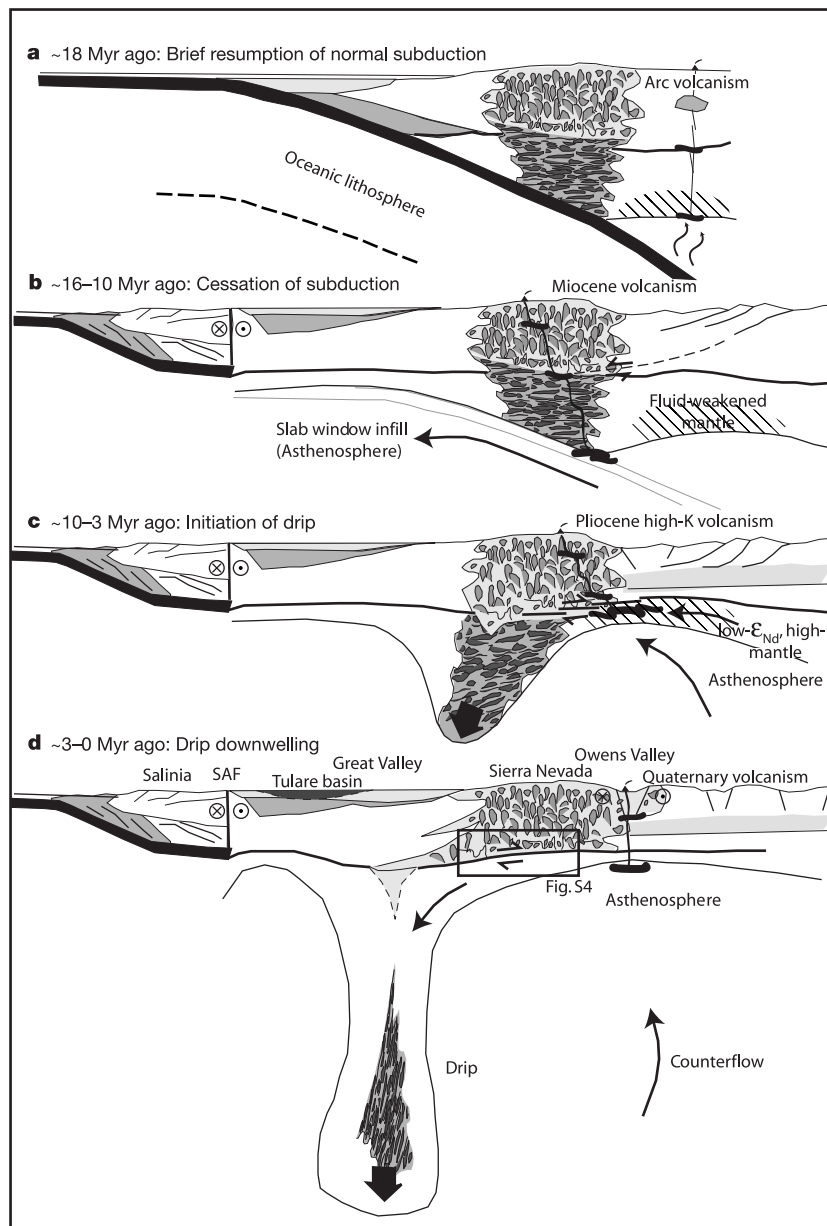


Figure 4 A sequential history of the foundering of the ultramafic root of the southern Sierra Nevada batholith. The proposed sequence is discussed in the text.

align in a direction antithetic to the shear direction; hence such features would be consistent with bottom-to-the-west shear. Laboratory experiments on feldspars at high temperatures and large shear strains showed melt segregated into melt-rich bands oriented obliquely to the shear plane and antithetic to the shear deformation³³. Fluids or melts are generally required to open and propagate fractures at these depths. (We know that at least some existed because even the limited Sierra Nevada volcanism³⁴ brought up xenoliths from such depths.) This interpretation requires high temperatures to trigger dehydration reactions or melting to induce fracturing. Although low temperatures are usually associated with convective instabilities³⁵, the formation of a shear zone between heterogeneous media can produce significant viscous heating³⁶. Even without shear heating, the rapid thinning of the lithosphere associated with asymmetric instabilities may be sufficient to locally raise mantle temperatures significantly. Quaternary volcanism, xenolith geothermometry and magnetotelluric measurements suggest the relatively recent attainment of high Moho temperatures (~1,000 °C), at least beneath Owens Valley^{7,8,37}. In an alternative interpretation that the anisotropy is due to oriented micas, the fabric is probably controlled by the development of an S-C fabric. In this case, the observed orientation of the anisotropy indicates top-to-the-west shear³², the opposite sense of the root moving westward into the drip (top-to-the-east sense of shear) but possibly consistent with an earlier phase of extension, suggesting a two-stage evolution involving a reversal in the sense of slip (see Supplementary Information).

Destabilization, detachment and sinking

Combining the new seismic results with the existing geologic constraints, we suggest the following sequence of events in the southern Sierra Nevada (Fig. 4). A dense root originated as an ultramafic phase of Sierran batholith formation during Mesozoic subduction. The Cretaceous granitic batholith and its garnet- and pyroxene-rich root formed a strong microplate under the influence of a cool geotherm that initiated during Laramide flat-slab subduction and is still reflected in modern heat flow (see, for example, ref. 38) and geothermometry (for example, refs 6, 39, 40). A number of factors could have played a role in destabilizing the root in the early Miocene. First, the volcanic arc returned to the northern Sierra (Fig. 4a), just within and north of the area pictured in Fig. 1 (ref. 41), terminating any unusually cool temperatures at the base of the continental lithosphere and possibly fluxing the Sierran lithosphere. Second, the Mendocino triple junction passed through this region between 10 and 20 Myr ago, exposing the Sierran lithosphere to the asthenosphere of the 'slab window' (Fig. 4b)⁴² and changing the basal stresses on the root¹¹. Third, extensional faults rooted into the Sierran crust developed starting at ~15 Myr ago (see, for example, ref. 34), weakening the top of the ultramafic root⁴³.

Although we have no direct information on the critical initial stages of the removal process, we suggest a history consistent with numerical models of lithospheric deformation in which the mechanical lithosphere (including the root) deforms by localized strain in the upper region in conjunction with Rayleigh–Taylor-type viscous deformation in the lower portion³⁰ (Fig. 4c). As the instability grew, a pronounced asymmetry in the foundering developed, resulting in a concentrated downwelling under the San Joaquin Valley and western foothills of the Sierra Nevada accompanied by a more widespread mantle upwelling under the western Basin and Range and easternmost Sierra (Fig. 4d). During this latter stage of development, the downwelling created a region of strong convergence in the mantle while the crust remained in a state of extension, suggesting decoupling (or a weak shear zone) between the crust and mantle except where the flow turns downward.

One explanation for the westward location of the downwelling is that the convective instability was initiated on the western edge of the batholith owing to compositional variations¹⁰, but a change in

the initiation location alone does not explain the asymmetric nature of the ensuing instability. Another explanation is that the instability was initially centred under the main crest of the Sierra but was displaced westward by asthenospheric flow related to the mantle wind⁴⁴ or to westward influx of asthenosphere into the slab window, a flow direction controlled by the shape and polarity of the palaeo-subduction zone^{45,46}. A third possibility, suggested by recent tomographic work⁴⁷, is that the garnet-rich material has descended roughly vertically but the removal has propagated from east to west, producing a dense, east- to southeast-dipping body connected at present to the crust at the Moho hole.

Numerical modelling of detachment of a lithospheric root that incorporates brittle crustal rheologies demonstrates that asymmetric formation of a convective instability is possible only if a prescribed dipping shear zone³⁰ or a laterally limited zone of weak brittle rheology is present²⁹. The hydrated lithosphere just east of the Sierra Nevada presumably established in the early Cenozoic above the subducting Farallon plate⁴⁸ is a likely location for such a laterally limited weak zone, especially given the initiation of extension in this region starting at ~15 Myr ago (see, for example, ref. 34). Westward motion of the detaching root would draw in weakened North American-affinity mantle from beneath the adjacent Basin and Range. Small-volume partial melting of this lithosphere mantle can reasonably explain the high-potassium, low- ϵ_{Nd} volcanism that occurred in the Sierra Nevada during the Pliocene¹⁵. Another important consequence of the asymmetric geometry of the removal process is the implication that the separation of the batholith from its root occurred along a concentrated zone of simple shear at the present-day Moho.

An important question remains unanswered: How much of the North American Cordilleran magmatic arc developed and then lost an ultramafic root? It is still debated whether the entire Sierra Nevada formed and lost its root^{16,44} or whether the removal is progressing northward and only the southern Sierra root is sinking, with the root remaining intact or in a nascent drip stage under the northern Sierra Nevada. The tectonic environment is obviously important in determining whether and when root removal occurs, but more documented examples are needed to clarify these relationships. Although not the focus of this Article, root foundering has important tectonic and geomorphic consequences—for example, collapse of the Salinian–Mojave block⁴⁹, uplift of the Sierra Nevada and its influence on adjacent ranges and faults¹⁶, and localized subsidence²⁴. Root removal is an important tectonic process with global implications, not only in terms of long-term crustal compositional evolution, but also in providing insights on lower-crustal deformation mechanisms, continental strength, and influence on regional tectonics. □

Received 16 February; accepted 12 July 2004; doi:10.1038/nature02847.

1. Ducea, M. N. The California arc: Thick granitic batholiths, eclogitic residues, lithospheric-scale thrusting, and magmatic flare-ups. *GSA Today* **11**, 4–10 (2001).
2. Kay, R. W. & Kay, S. M. Creation and destruction of lower continental crust. *Geol. Rundsch.* **80**, 259–278 (1991).
3. Kay, R. W. & Kay, S. M. Delamination and delamination magmatism. *Tectonophysics* **219**, 177–189 (1993).
4. Ducea, M. N. Constraints on the bulk composition and root foundering rates of continental arcs: A California arc perspective. *J. Geophys. Res.* **B 107**, doi:10.1029/2001JB000643 (2002).
5. Jull, M. & Kelemen, P. B. On the conditions for lower crustal convective instability. *J. Geophys. Res.* **106**, 6423–6446 (2001).
6. Ducea, M. N. & Saleeby, J. B. Buoyancy sources for a large, unrooted mountain range, the Sierra Nevada, California: Evidence from xenolith thermobarometry. *J. Geophys. Res.* **101**, 8229–8244 (1996).
7. Ducea, M. N. & Saleeby, J. B. A case for delamination of the deep batholithic crust beneath the Sierra Nevada, California. *Int. Geol. Rev.* **40**, 78–93 (1998).
8. Ducea, M. N. & Saleeby, J. B. The age and origin of a thick mafic-ultramafic keel from beneath the Sierra Nevada batholith. *Contrib. Mineral. Petrol.* **133**, 169–185 (1998).
9. Wernicke, B. et al. Origin of high mountains in the continents: The Southern Sierra Nevada. *Science* **271**, 190–193 (1996).
10. Saleeby, J., Ducea, M. & Clemens-Knott, D. Production and loss of high-density batholithic root—Sierra Nevada, California. *Tectonics* **22**, doi:10.1029/2002TC001374 (2003).
11. Zandt, G. & Carrigan, C. R. Small-scale convective instability and upper mantle viscosity under California. *Science* **261**, 460–463 (1993).
12. Jones, C. H., Kanamori, H. & Roecker, S. W. Missing roots and mantle "drips": Regional Pn and

- teleseismic arrival times in the southern Sierra Nevada and vicinity, California. *J. Geophys. Res.* **99**, 4567–4601 (1994).
13. Ruppert, S., Fliedner, M. & Zandt, G. Thin crust and active upper mantle beneath the Southern Sierra Nevada in the western United States. *Tectonophysics* **286**, 237–252 (1998).
14. Manley, C. R., Glazner, A. F. & Farmer, G. L. Timing of volcanism in the Sierra Nevada of California: Evidence for Pliocene delamination of the batholithic root? *Geology* **28**, 811–814 (2000).
15. Farmer, G. L., Glazner, A. F. & Manley, C. Did lithospheric delamination trigger late Cenozoic potassic volcanism in the Sierra Nevada, California? *Geol. Soc. Am. Bull.* **114**, 754–768 (2002).
16. Jones, C. H., Farmer, G. L. & Unruh, J. Tectonics of Pliocene delamination of lithosphere of the Sierra Nevada, California. *Geol. Soc. Am. Bull.* (in the press).
17. Dueker, K. G. & Sheehan, A. F. Mantle discontinuity structure from mid-point stacks of converted P to S waves across the Yellowstone hotspot track. *J. Geophys. Res.* **102**, 8313–8327 (1997).
18. Kind, R. *et al.* Comprehensive seismic images of the crust and upper mantle beneath Tibet. *Science* **298**, 1219–1222 (2002).
19. Gilbert, H. J., Sheehan, A. F., Dueker, K. G. & Molnar, P. Receiver functions in the western United States, with implications for upper mantle structure and dynamics. *J. Geophys. Res.* **B 108**, doi:10.1029/2001JB001194 (2003).
20. Fliedner, M., Klemperer, S. L. & Christensen, N. I. Three-dimensional seismic model of the Sierra Nevada arc, California, and its implications for crustal and upper mantle composition. *J. Geophys. Res.* **105**, 10899–10921 (2000).
21. Zhu, L. & Kanamori, H. Moho depth variation in southern California from teleseismic receiver functions. *J. Geophys. Res.* **105**, 2969–2980 (2000).
22. Bostock, M. G., Hyndman, R. D., Rondenay, S. & Peacock, S. M. An inverted continental Moho and serpentinization of the forearc mantle. *Nature* **417**, 536–538 (2002).
23. Jones, C. H. & Phinney, R. A. Seismic structure of the lithosphere from teleseismic converted arrivals observed at small arrays in the southern Sierra Nevada and vicinity, California. *J. Geophys. Res.* **103**, 10065–10090 (1998).
24. Saleeby, J. & Foster, Z. Topographic response to mantle lithosphere removal in the southern Sierra Nevada region, California. *Geology* **37**, 245–248 (2004).
25. Liu, M. & Shen, Y. Q. Sierra Nevada uplift: A ductile link to mantle upwelling under the basin and range province. *Geology* **26**, 299–302 (1998).
26. Furlong, K. P. & Govers, R. Ephemeral crustal thickening at a triple junction: The Mendocino crustal conveyor. *Geology* **27**, 127–130 (1998).
27. Neil, E. A. & Houseman, G. A. Rayleigh–Taylor instability of the upper mantle and its role in intraplate orogeny. *Geophys. J. Int.* **138**, 89–107 (1999).
28. Houseman, G., Neil, E. A. & Kohler, M. D. Lithospheric instability beneath the Transverse Ranges of California. *J. Geophys. Res.* **105**, 16237–16250 (2000).
29. Schott, B. & Schmeling, H. Delamination and detachment of a lithospheric root. *Tectonophysics* **296**, 225–247 (1998).
30. Pysklywec, R. N., Beaumont, C. & Fullsack, P. Lithospheric deformation during the early stages of continental collision: Numerical experiments and comparison with South Island, New Zealand. *J. Geophys. Res.* **B 107**, doi:10.1029/2001JB000252 (2002).
31. Bindschadler, D. L. & Parmentier, E. M. Mantle flow tectonics: The influence of a ductile lower crust and implications for the formation of topographic uplands on Venus. *J. Geophys. Res.* **95**, 21329–21344 (1990).
32. Davis, G. H. & Reynolds, S. J. *Structural Geology of Rocks and Regions* 2nd edn (Wiley & Sons, New York, 1996).
33. Zimmerman, M. E. & Kohlstedt, D. L. Melt segregation and LPO in anorthite-basalt deformed in torsion. *Eos* **84** (Fall Meet. Suppl.), S22E–06 (2003).
34. Wernicke, B. P. in *The Geology of North America* Vol. G-3, *The Cordilleran Orogen: Conterminous US* (eds Burchfiel, B. C., Lipman, P. W. & Zoback, M. L.) 553–581 (Geological Society of America, Boulder, Colorado, 1992).
35. Houseman, G. A. & Molnar, P. Gravitational (Rayleigh–Taylor) instability of a layer with non-linear viscosity and convective thinning of continental lithosphere. *J. Geophys. Res.* **128**, 125–150 (1997).
36. Schott, B., Yuen, D. A. & Schmeling, H. Viscous heating in heterogeneous media as applied to the thermal interaction between crust and mantle. *Geophys. Res. Lett.* **26**, 513–516 (1999).
37. Park, S. K., Hirasuna, B., Jiracek, G. R. & Kinn, C. L. Magnetotelluric evidence of lithospheric mantle thinning beneath the southern Sierra Nevada. *J. Geophys. Res.* **101**, 16241–16255 (1996).
38. Saltus, R. W. & Lachenbruch, A. H. Thermal evolution of the Sierra Nevada: Tectonic implications of new heat flow data. *Tectonics* **10**, 325–344 (1991).
39. Dumitru, T. A. Subnormal Cenozoic geothermal gradients in the extinct Sierra Nevada magmatic arc: Consequences of Laramide and post-Laramide shallow-angle subduction. *J. Geophys. Res.* **95**, 4925–4942 (1990).
40. House, M. A., Farley, K. F., Wernicke, B. P. & Dumitru, T. A. Cenozoic thermal evolution of the central Sierra Nevada from (U–Th)/He thermochronology. *Earth Planet. Sci. Lett.* **151**, 167–179 (1997).
41. Dickinson, W. R. The Basin and Range Province as a composite extensional domain. *Int. Geol. Rev.* **44**, 1–38 (2002).
42. Atwater, T. & Stock, J. Pacific–North America plate tectonics of the Neogene southwestern United States: An update. *Int. Geol. Rev.* **40**, 375–402 (1998).
43. Molnar, P. & Jones, C. H. A test of laboratory-based rheological parameters of olivine from an analysis of late Cenozoic convective removal of mantle lithosphere beneath the Sierra Nevada, California, USA. *Geophys. J. Int.* **156**(3), 555–564 (2004).
44. Zandt, G. The southern Sierra Nevada drip and the mantle wind direction beneath the southwestern United States. *Int. Geol. Rev.* **45**, 213–224 (2003).
45. Furlong, K. P., Lock, J., Guzowski, C., Whitlock, J. & Benz, H. in *The Lithosphere of Western North America and its Geophysical Characterization, The George A. Thompson Volume* (eds Klemperer, S. L. & Ernst, W. G.) 92–104 (International Book Series, Vol. 7, Bellwether Publishing/Geological Society of America, Boulder, Colorado, 2003).
46. Hartog, R. & Schwartz, S. Y. Subduction induced strain in the upper mantle east of the Mendocino triple junction, California. *J. Geophys. Res.* **105**, 7909–7930 (2000).
47. Boyd, O. S., Jones, C. H. & Sheehan, A. F. Foundering lithosphere imaged beneath the southern Sierra Nevada, California, USA. *Science* **305**, 660–662 (2004).
48. Lange, R. A., Carmichael, I. S. E. & Renne, P. R. Potassic volcanism near the Mono basin, California: Evidence for high water and oxygen fugacities inherited from subduction. *Geology* **21**, 949–952 (1993).
49. Saleeby, J. B. Segmentation of the Laramide slab—Evidence from the southern Sierra Nevada region. *Geol. Soc. Am. Bull.* **115**, 655–668 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements G.Z. thanks G. Gehrels, P. Kapp and B. Hacker for comments on preliminary interpretations and manuscripts.

Authors' contributions G.Z., H.G., T.J.O. and C.H.J. cooperated on the seismology analysis and interpretation. M.D. and J.S. provided the geologic and tectonic context. C.H.J. led the PASSCAL deployment to collect the data. G.Z. wrote the Article with contributions from all the authors.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.Z. (zandt@geo.arizona.edu).

Inscribed matter as an energy-efficient means of communication with an extraterrestrial civilization

Christopher Rose¹ & Gregory Wright²

¹WINLAB, Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854, USA

²Antiope Associates, 18 Clay Street, Fair Haven, New Jersey 07704, USA

It is well known that electromagnetic radiation—radio waves—can in principle be used to communicate over interstellar distances^{1,2}. By contrast, sending physical artefacts has seemed extravagantly wasteful of energy, and imagining human travel between the stars even more so^{3,4}. The key consideration in earlier work, however, was the perceived need for haste. If extraterrestrial civilizations existed within a few tens of light years, radio could be used for two-way communication on timescales comparable to human lifetimes (or at least the longevity of human institutions). Here we show that if haste is unimportant, sending messages inscribed on some material can be strikingly more energy efficient than communicating by electromagnetic waves. Because messages require protection from cosmic radiation and small messages could be difficult to find among the material clutter near a recipient, ‘inscribed matter’ is most effective for long archival messages (as opposed to potentially short “we exist” announcements). The results suggest that our initial contact with extraterrestrial civilizations may be more likely to occur through physical artefacts—essentially messages in a bottle—than via electromagnetic communication.

We consider a message of B bits to be sent over a distance D and received by time τ . For inscribed matter we assume the destination is at rest relative to the source, that the mass packet is accelerated with a launcher (as opposed to an on-board engine) and for the moment, that the packet is not decelerated but is ‘caught’ at its destination. For radiation, the transmission is of duration T , so the entire message is available at the receiver after a delay of $\tau = D/c + T$. We use message delivery energy as a proxy for cost since it represents a required minimum of resources, independent of technological or organizational skill.

The energy required to deliver inscribed matter a distance D by time τ in free space is minimized if the particle is launched at speed $v = D/\tau$. We parameterize acceptable delay by a dimensionless quantity $\delta = c\tau/D$ where $\delta \geq 1$ means that we require the message to be available at a time just greater than the light transit delay, and $\delta \gg 1$ means we can tolerate a long delay. For a message of size B using matter with mass information density of $\bar{\rho}$ (bits kg^{-1}) we then have energy:

$$E_w = \frac{1}{2} \left[\frac{B}{\bar{\rho}} \right] \left(\frac{c}{\delta} \right)^2 \quad (1)$$

Notice that we assume messages travel at non-relativistic speeds. Relativistic effects are only important when $v \geq 0.7c$, and by our original assumption, we are willing to accept delays much greater than the light transit time. This point is worth emphasizing, because a perceived need for haste led to the conclusion that delivery of matter over interstellar distances is energetically prohibitive^{3,4}. We also note that particle delivery might require overcoming a gravitational potential, but carefully computing such penalties does not greatly affect our main points.

To send the same message of B bits between antennas of equal aperture R separated by distance D requires energy:

$$E_r \geq BN_0 \frac{\lambda^2 D^2}{2\pi^2 R^4} \ln 2 \quad (2)$$

with operating wavelength λ and noise spectral density N_0 . This lower bound is derived from information theory and assumes diffraction-limited optics^{5,6}. This is a best-case scenario for electromagnetic communication that sidesteps the admittedly interesting questions of preferred frequencies and bandwidths considered by others^{1,2,7}. Our energy estimate is conservative because no method of electromagnetic communication can use less energy than is given by equation (2). A full derivation can be found in the Supplementary Information.

We can now define Ω , the radiation-to-inscribed-matter energy ratio, as:

$$\Omega = \frac{E_r}{E_w} \geq \frac{16 \ln 2}{\pi^2} \left[\frac{\bar{\rho} N_0}{c^2} \right] \left[\frac{\mathcal{D}}{\mathcal{A}} \right]^2 \delta^2 \quad (3)$$

where $\mathcal{A} = 2R/\lambda$ is the normalized antenna aperture and $\mathcal{D} = D/2R$ is the distance between transmitter and receiver in units of antenna aperture. This is our main result. $\Omega > 1$ implies more energy per bit to radiate information than to deliver it via inscribed matter. As expected, Ω increases with $\mathcal{D}$ owing to the unavoidable dispersal of electromagnetic radiation with distance, and decreases with the better collimation afforded by increased antenna aperture $\mathcal{A}$. We also see that relaxing the requirement of haste benefits inscribed matter by a factor δ^2 .

How well inscribed matter performs depends on how densely information can be written. A scanning tunnelling microscope (STM) can place an equivalent of about 10^{15} bits per square inch using individual xenon atoms on a nickel substrate⁸. The per bit dimension is then 0.8 nm on a side. Assuming a 10-nm nickel buffer between layers, we obtain a bit density of 1.56×10^{20} bits cm^{-3} , or $\bar{\rho}_{\text{STM}} = 1.8 \times 10^{22}$ bits kg^{-1} . If we could build stable alloys of the lightest solid elements (Li, Be) with arbitrary placement of the atoms, we could achieve $\bar{\rho}_{\text{LiBe}} = 7.5 \times 10^{25}$ bits kg^{-1} . For comparison, the information density of single-stranded RNA (for example, polio virus RNA) is about 3.6×10^{24} bits kg^{-1} . We adopt a conservative $\bar{\rho} = 10^{22}$ bits kg^{-1} (equivalent to about 1000 nickel atoms per bit) as our reference mass information density.

Message assembly is part of the inscribed-matter energy budget, and even if in theory this assembly energy is zero⁹, it is useful to have empirical bounds. As a worst case, message assembly might require construction from individual atoms. Assuming an energy of about 2 eV per bond leads to about 2,000 eV bit^{-1} for assembly of a nickel-based message. Using Solar System escape velocity from Earth orbit (42 km s^{-1}) and $\bar{\rho} = 10^{22}$ bits kg^{-1} , we find that the launch energy per bit is 5.5×10^5 eV bit^{-1} or 275 times the 2,000 eV bit^{-1} assembly energy. We therefore expect inscription energy cost to be negligible relative to launch energy cost at the speeds required for interstellar delivery. A more complete discussion of these issues can be found in the Supplementary Information.

There are two more potentially serious energy penalties for inscribed matter. The message must be either encoded or shielded to preclude irreparable radiation damage during its journey, and if the message is to await discovery at its destination, some means of deceleration is required for gravitational capture by the target.

To reach a star ten thousand light years away at a speed of $10^{-3}c$, an inscribed-matter packet would endure a trip of ten million years and a variety of insults from the interstellar medium. If, for example, 99% of the bits in a message were randomly erased by cosmic rays, the coding needed to ensure error-free reception requires us to send 100 bits for every bit in the original message⁵. If messages are more fragile, we can also consider shielding. Recent studies¹⁰ show that 800 g cm^{-2} of shielding material could support 10% viability of bacterial spores (*Deinococcus radiodurans*) subject to cosmic radiation for approximately 30 million years. Assuming shield material with an approximately terrestrial average density of 3 g cm^{-3} , 10% survival of a 1 kg information payload incurs a total mass to (surviving) information penalty of roughly 2.4×10^6 .

An on-board means of decelerating a message of size B at the destination reduces the efficiency by a factor of $\exp(c/\delta g I_{\text{sp}})$, because mass must be expelled as exhaust to brake the craft. The product $g I_{\text{sp}}$, where g is the Earth gravitational acceleration, is just the exhaust velocity of the braking rocket. This parameterization, introducing the specific impulse I_{sp} , is conventional for comparing the efficiencies of different rockets. Values of I_{sp} range from hundreds of seconds for chemical engines to 10^6 seconds for antimatter annihilation engines^{11,12}. For $I_{\text{sp}} = 2 \times 10^4$ (nuclear electric engine) and $\delta = 1,000$, the braking penalty factor $\exp(c/\delta g I_{\text{sp}})$ would be 4.6.

In Fig. 1 we provide iso- Ω contours as a function of $\mathcal{D}$ and $\mathcal{A}$ for communications with $\delta = 10^3$ (about seven times larger than solar escape velocity at Earth distance) and $\bar{\rho} = 10^{22}$ bits kg^{-1} . Points corresponding to examples of receiver apertures and operating wavelengths at various distances are shown. We see that inscribed matter is energetically favoured over a wide range of conditions. For instance, inscribed matter is more efficient than radiation between Arecibo-sized apertures for distances greater than 1.32×10^{12} m ($\mathcal{D} = 4.4 \times 10^9$), about the distance from the Sun to Saturn. Likewise, 10-m optical and 1-m X-ray systems are less efficient than inscribed matter for distances beyond about 0.04 and 2 light years, respectively.

As a concrete example, consider the Voyager spacecraft carrying an analogue recording containing about 10^9 bits of information. The spacecraft has a mass of about 10^3 kg. If Voyager had been fired from a catapult rather than a rocket, its launch energy of 800 J bit^{-1} would make it more efficient than Arecibo-to-Arecibo radio communication for distances beyond about 2,000 light years. Were Voyager carrying three DVD disks (about 10^{11} bits), this 'break-even

distance' would be 200 light years. With 100 g of $\bar{\rho} = 10^{22}$ material, break-even would occur at 2×10^{-3} light years. And if rocket were used rather than a catapult, all break-even distances would be a factor of about nine larger.

For a 10^4 light year trip, the combined penalty for conservative shielding and deceleration is approximately 10^7 , but this is insufficient to negate the very large inscribed-matter advantage over radiation seen in Fig. 1. We have also not accounted for the fact that a radio message must be repeated many times to have a high probability of being received by a destination that may only occasionally be listening. Accounting for this effect would make inscribed matter orders of magnitude more attractive, as shown in the Supplementary Information.

Having shown that writing can use much less energy than radiating leaves us with two practical questions: does sending written messages across interstellar space really confer a practical advantage over radio, and if so, what are the implications for extraterrestrial message search from Earth?

Could the total cost of an effort to communicate across interstellar distances be less using inscribed matter? Equation (3) is a bound on marginal efficiency. That is, Ω is the minimum energy needed to send an additional bit, ignoring the cost of building the infrastructure to send messages by either radiation or inscribed matter. To benefit from the efficiencies promised by equation (3), a civilization must be able to launch messages that can navigate between the stars. It would seem that a simple radio beacon would require fewer resources and thus might be the preferred means for a civilization to announce itself.

Although narrow-band radio beacons could be efficient for short "we exist" messages, if a civilization wants to send more than just a few bits of information a substantial distance across the Galaxy, then it is constrained by physics to build very large antennas—as large as the Earth if using microwave frequencies—or to consider sending a message inscribed on matter. Although we cannot know the actual level of resources needed to build an interstellar communication system, we can say that if long messages are desired, the irreducible expenditure of energy reflected by equation (3) becomes more important and favours inscribed matter.

If interstellar inscribed-matter messages are indeed more efficient, then it is natural to ask where they might be found in our Solar System. Messages designed to await discovery would have to remain in orbits of long-term gravitational stability, or on the surface of objects in such orbits. The stable Lagrange points (L4 and L5) of Jupiter and the Sun, L4 and L5 of the Earth and the Moon, orbits close to the Sun, low-eccentricity orbits in the main asteroid belt¹³ and perhaps similar orbits in the Kuiper belt offer such zones of dynamical refuge. The surfaces of various bodies in the inner Solar System are also possibilities.

Any message presumably arrived after the Solar System became habitable (that is, after most of the protoplanetary debris had cleared), so whatever carried the message would be less eroded by impacts than an asteroid. Interplanetary radar could search for objects with anomalously smooth radar signatures. Alternatively, a message could have a retroreflector attached to produce an anomalously large radar cross-section. Of course, an even simpler strategy is to use a powerful radio beam to illuminate these regions and see whether anything answers back. More-active message types (ecological seeds¹⁴ or probes¹⁵) are also conceivable, but are not necessary to exploit inscribed-matter efficiency.

Our results suggest that carefully searching our own planetary backyard may be as likely to reveal evidence of extraterrestrial civilizations as studying distant stars through telescopes. □

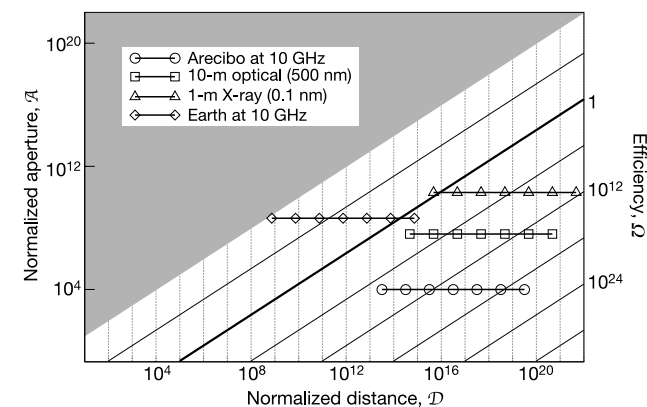


Figure 1 Efficiency of inscribed matter compared to electromagnetic communication. Contours of constant Ω are plotted as a function of $\mathcal{A} = 2R/\lambda$ (a measure of transmit antenna directivity) and $\mathcal{D} = D/2R$ (a measure of receiver power capture) using equation (3). We assume equal-sized transmit and receive apertures, receiver noise temperature 3 K, $\delta = 10^3$ and $\bar{\rho} = 10^{22}$ bits kg^{-1} . Below and to the right of the line labelled $\Omega = 1$, inscribed mass is a more energetically efficient means of communication. In the shaded region, aperture and distance are such that no additional advantage is conferred by tighter beam collimation—all radiated energy is captured at the target. In this parameterization, isotropic transmission is given by $\mathcal{A} = 1/\pi\sqrt{2}$, the smallest value of ordinate plotted. ($\mathcal{D}, \mathcal{A}$) points are also shown for various antenna aperture sizes (Arecibo, $R = 150$ m, an antenna the diameter of the Earth, $R = 6.38 \times 10^6$ m, 10-m optical and 1-m X-ray telescopes) and radiation wavelengths (0.03 m (10 GHz), 500 nm (optical) and 0.1 nm (X-ray)). Target ranges from 1 to 10^9 light years in one-light-year increments are depicted for each case. For reference we note that $\delta = 7,143$ at solar escape velocity near Earth (42 km s^{-1}), whereas here we use a speedier $\delta = 10^3$. At the high gains needed for radiation to be competitive with inscribed matter, radio beams do not illuminate many stars, vitiating any 'broadcast advantage' of radiation. A 10-GHz Arecibo-sized aperture pointed in the plane of our galaxy illuminates about 10^3 stars within 10^4 light years—a small incursion on the inscribed-matter advantage of about 5×10^{15} . Similarly oriented 10-m optical and 1-m X-ray aperture beams are narrow and enjoy no broadcast advantage out to 10^4 light years.

Received 8 April; accepted 16 July 2004; doi:10.1038/nature02884.

1. Cocconi, G. & Morrison, P. Searching for interstellar communications. *Nature* **184**, 844–846 (1959).
2. Schwartz, R. N. & Townes, C. H. Interstellar and interplanetary communication by optical masers. *Nature* **190** (4772), 205–208 (1961).

3. Purcell, E. W. Radio astronomy and communication through space. In *Interstellar Communication* Ch. 13 (ed. Cameron, A. G. W.) 121–143 (Benjamin, New York, 1963).
4. Pierce, J. R. Relativity and space travel. *Proc. IRE* **47**, 1053–1061 (1959).
5. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* 247–250, 187–189 (Wiley-Interscience, New York, 1991).
6. Skolnik, M. *Introduction to Radar Systems* 6.2–6.5 (McGraw-Hill, New York, 2002).
7. Townes, C. H. At what wavelengths should we search for signals from extraterrestrial intelligence? *Proc. Natl Acad Sci.* **80**, 1147–1151 (1983).
8. Eigler, D. M. & Schweizer, E. K. Positioning single atoms with a scanning tunnelling microscope. *Nature* **344**, 524–526 (1990).
9. Landauer, R. Minimal energy requirements in communication. *Science* **272**(5270), 1914–1918 (1996).
10. Mileikowsky, C. F. *et al.* Natural transfer of viable microbes in space: from Mars to Earth and Earth to Mars. *Icarus* **145**, 391–427 (2000).
11. Bussard, R. W. & DeLauer, R. D. *Fundamentals of Nuclear Flight* 16–21 (McGraw-Hill, New York, 1965).
12. Stuhlinger, E. *Ion Propulsion for Space Flight* (McGraw-Hill, New York, 1964).
13. Papagiannis, M. Are we alone or could they be in the asteroid belt? *Q. J. R. Astron. Soc.* **19**, 277–281 (1978).
14. Crick, F. H. C. & Orgel, L. E. Directed panspermia. *Icarus* **19**, 341–346, (1973).
15. Bracewell, R. Communication for superior galactic communities. *Nature* **187**, 670–671 (1960).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank G. Foschini, G. Gonzalez, P. Herry, P. Morrison, P. Shor and S. Tournpis for their comments on an earlier version of the manuscript. This work was supported, in part, by the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.R. (crose@winlab.rutgers.edu.).

Spectroscopy of spontaneous spin noise as a probe of spin dynamics and magnetic resonance

S. A. Crooker¹, D. G. Rickel¹, A. V. Balatsky² & D. L. Smith²

¹National High Magnetic Field Laboratory, ²Theory Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Not all noise in experimental measurements is unwelcome. Certain fundamental noise sources contain valuable information about the system itself—a notable example being the inherent voltage fluctuations (Johnson noise) that exist across any resistor, which allow the temperature to be determined^{1,2}. In magnetic systems, fundamental noise can exist in the form of random spin fluctuations^{3,4}. For example, statistical fluctuations of N paramagnetic spins should generate measurable noise of order $\sqrt{N}$ spins, even in zero magnetic field^{5,6}. Here we exploit this effect to perform perturbation-free magnetic resonance. We use off-resonant Faraday rotation to passively^{7,8} detect the magnetization noise in an equilibrium ensemble of paramagnetic alkali atoms; the random fluctuations generate spontaneous spin coherences that precess and decay with the same characteristic energy and timescales as the macroscopic magnetization of an intentionally polarized or driven ensemble. Correlation spectra of the measured spin noise reveal g -factors, nuclear spin, isotope abundance ratios, hyperfine splittings, nuclear moments and spin coherence lifetimes—without having to excite, optically pump or otherwise drive the system away from thermal equilibrium. These noise signatures scale inversely with interaction volume, suggesting a possible route towards non-perturbative, sourceless magnetic resonance of small systems.

The fluctuation–dissipation theorem states that the response of a system to an external perturbation (that is, the susceptibility) can be

described by the spectrum of fluctuations exhibited by the system in thermal equilibrium⁹. In magnetic systems, $\sqrt{N}$ fluctuations in an ensemble of N undriven nuclear spins has been observed⁶, as predicted by Bloch⁵ in 1946. Fundamental magnetic fluctuations of thermal (and quantum) origin have since been identified in, for example, spin glasses¹⁰, hard-disk magnetoresistive heads¹¹, and in the magnetic noise spectrum of antiferromagnetic particles¹². Optical techniques have identified the presence of stochastic spin fluctuations in atomic systems^{4,13–16}, and notably, fluctuations corresponding to very few spins—and perhaps even single spins—are evidenced by ultrasensitive cantilevers¹⁷ and scanning tunnelling microscopes^{18,19}, respectively. In other disciplines, thermal noise in nanomechanical resonators²⁰ and recent correlation spectra of thermal acoustic vibrations suggest a means for ‘sourceless ultrasonics’²¹. Here we investigate the detailed spectroscopy of spin noise to perform perturbation-free magnetic resonance. We study ground-state magnetization fluctuations in a classical ensemble of uncorrelated paramagnetic spins in thermal equilibrium, realized in atomic alkali vapours. Random spin fluctuations and their associated coherences reveal the complete magnetic structure of the atomic $^2S_{1/2}$ ground state, including hyperfine, Zeeman and nuclear moment effects. Historically, this information is obtained with conventional magnetic resonance techniques (optical pumping and/or radio-frequency excitation)^{22–24}, which necessarily perturb the spin ensemble away from thermal equilibrium.

Figure 1a shows a diagram of our experiment. A Ti:sapphire ring laser (~ 8 GHz linewidth), detuned from any atomic absorption, is linearly polarized and focused through a cell containing a

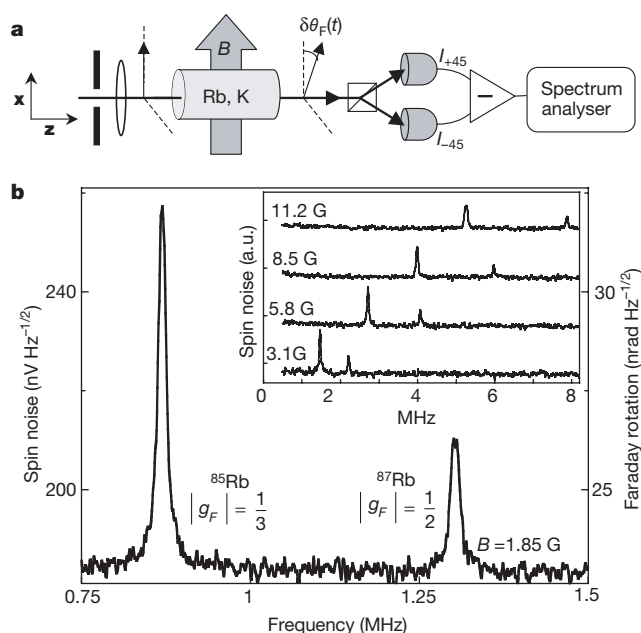


Figure 1 Spontaneous spin noise in Rb or K vapour, probed via Faraday rotation.

a, Experimental schematic. Ground-state stochastic spin fluctuations $\delta M_z(t)$ impart polarization fluctuations $\delta \theta_F(t)$ on the detuned probe laser. **b**, Measured spectrum of spin (Faraday rotation) noise from Rb vapour at temperature $T = 369$ K and magnetic field $B = 1.85$ G, showing spontaneous spin coherence peaks from ^{85}Rb and ^{87}Rb . The data are given in units of root-mean-square (r.m.s.) spectral density of voltage fluctuations ($\text{nV Hz}^{-1/2}$) measured in the photodiode bridge, and also by the r.m.s. spectral density of Faraday rotation fluctuations ($\text{nrad Hz}^{-1/2}$). The noise floor is determined primarily by photon shot noise. The laser is detuned $\Delta_{D1} = 25$ GHz from the D1 transition ($5^2S_{1/2}$ to $5^2P_{1/2}$, ~ 794.8 nm), ensuring negligible absorption. Inset: the ^{85}Rb and ^{87}Rb spin noise peaks measured at $\Delta_{D2} = 20$ GHz from the D2 transition ($5^2S_{1/2}$ to $5^2P_{3/2}$, ~ 780 nm). The magnetic field B is incremented in steps of 2.7 G, and $T = 369$ K. Plots offset vertically for clarity. a.u., arbitrary units.

40-mm-long, temperature-controlled column of rubidium or potassium vapour having typical density of order 10^9 atoms mm^{-3} . Random magnetization fluctuations along z impart small polarization (Faraday) rotation fluctuations $\delta\theta_F(t)$ on the laser, which are measured with a balanced photodiode bridge. Helmholtz coils provide a small transverse magnetic field B along x , about which all magnetization fluctuations δM_z must precess. This precession, which corresponds to transverse spin coherences between Zeeman sublevels, shifts the spin fluctuation signal away from very low frequencies where environmental noise dominates. The detuned laser ensures a perturbation-free probe⁷ of equilibrium spin noise, wherein the $\sim 10^9$ atoms within the laser volume are not optically pumped or excited in any way. The ensemble exhibits zero net magnetization ($\langle M_z(t) \rangle = 0$), with a nominally equal population of spins within ground-state sublevels.

Near the D1 or D2 transition of alkali atoms, the indices of refraction for right- and left-circularly polarized light are $n^\pm - 1 \cong (N_0 e^2 f / 4\pi m \nu \Delta)(1 \pm \beta J_e)$, where N_0 is the density of atoms, f and β are the transition's oscillator strength and polarizability ($\beta = 2$ and -1 for D1 and D2, respectively), ν is the laser frequency, and $\Delta = \nu - \nu_0$ is the laser detuning from the transition centre (assumed here to be much larger than the pressure-broadened transition width). $\langle J_e \rangle \equiv (N_0^+ - N_0^-) / 2N_0$ is the expectation value of the valence electron spin along z , where $N_0^\pm$ are the densities of atoms with ground-state spin projection parallel and antiparallel to the laser. The number of atoms within the laser beam of cross-sectional area A and over length L is $N = N_0 AL$. Magnetization noise arises from statistical fluctuations in the quantity $N^+ - N^-$, which has amplitude $\sqrt{\langle (N^+ - N^-)^2 \rangle} = \sqrt{N} = \sqrt{N_0 AL}$. The induced Faraday rotation, $\theta_F = \pi \nu L(n^+ - n^-) / c$, therefore exhibits fluctuations:

$$\langle \delta\theta_F^2 \rangle^{1/2} = \frac{e^2 f \beta}{4\pi m c \Delta} \sqrt{\frac{N_0 L}{A}} \quad (1)$$

The spin correlation function, $S(t) = \langle M_z(0)M_z(t) \rangle$, has a Fourier transform $S(\omega)$ that is proportional to the measured power spectrum of $\delta\theta_F(t)$.

A typical noise spectrum from rubidium vapour is shown in Fig. 1b, taken with the laser detuned 25 GHz from the D1 transition. The sharp peaks at frequencies $\Omega = 869$ and $1,303$ kHz are due to random spin fluctuations that are precessing in the small 1.85 G transverse magnetic field, effectively generating spontaneous spin coherences between ground-state Zeeman sublevels. These coherences precess with effective g -factors $g_F = h\Omega / \mu_B B \approx 1/3$ and $1/2$, which are the ground-state g -factors of the stable isotopes ^{85}Rb and ^{87}Rb (h and μ_B are the Planck constant and Bohr magneton,

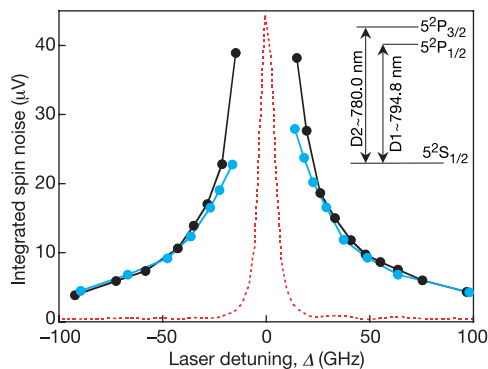


Figure 2 Total integrated spin noise from ^{85}Rb as a function of laser detuning from the D1 (black points) and D2 (blue points) transitions. The laser power was $200 \mu\text{W}$, $T = 359.1$ K, and the D1 (D2) transition was pressure-broadened with 125 torr (250 torr) of N_2 buffer gas. Lines are guides to the eye. For reference, the dotted red line is the measured D2 optical depth (~ 2.5 at peak, or $\sim 92\%$ absorption).

respectively). Coupling of the nuclear spin I to the $J = 1/2$ valence electron splits the $^2\text{S}_{1/2}$ atomic ground state into two hyperfine F -levels with total spin $F = I \pm J$ and g -factor $|g_F| \cong g_I / (2I + 1)$, where $g_I \cong 2$ is the free electron g -factor. Thus, the nuclear spin of ^{85}Rb ($I = 5/2$) and ^{87}Rb ($I = 3/2$) may be directly measured from spin fluctuations in thermal equilibrium. Noise spectra acquired near the D2 transition show similar peaks (inset, Fig. 1b), which move as expected with magnetic field. The 13 kHz measured width of these noise peaks indicates an effective transverse spin dephasing time of $\sim 100 \mu\text{s}$, much less than the known Rb spin lifetime (~ 1 s), due largely to the transit time of atoms across the $\sim 100 \mu\text{m}$ laser diameter. The spectral density of the spin noise is small—the ^{87}Rb peak in Fig. 1b contributes only $3.1 \text{ nrad Hz}^{-1/2}$ of Faraday rotation noise above the photon shot noise floor of $23 \text{ nrad Hz}^{-1/2}$. Because spin noise arises, effectively, from N uncorrelated precessing spins, the integrated area under the noise peaks should scale with $\sqrt{N}$. This is conveniently confirmed by noting that the integrated spin noise of the ^{85}Rb and ^{87}Rb peaks is 24.7 and $15.4 \mu\text{V}$ respectively, whose ratio— 1.60 —agrees quite well with the square root of the ^{85}Rb : ^{87}Rb natural abundance ratio ($\sqrt{72.2\% / 27.8\%} = 1.61$).

That the off-resonant laser non-perturbatively probes spin fluctuations is evidenced in Fig. 2, where the integrated ^{85}Rb spin noise is measured at large detuning from both the D1 and D2 transitions. The measured D2 absorption (optical depth) is also shown for reference. Appreciable spin noise exists out to $\Delta = \pm 100$ GHz, increasing at smaller detunings down to 15 GHz, at which point noticeable absorption occurs. For $|\Delta| > 20$ GHz, the measured spin noise scales as Δ^{-1} , confirming that the off-resonant laser is

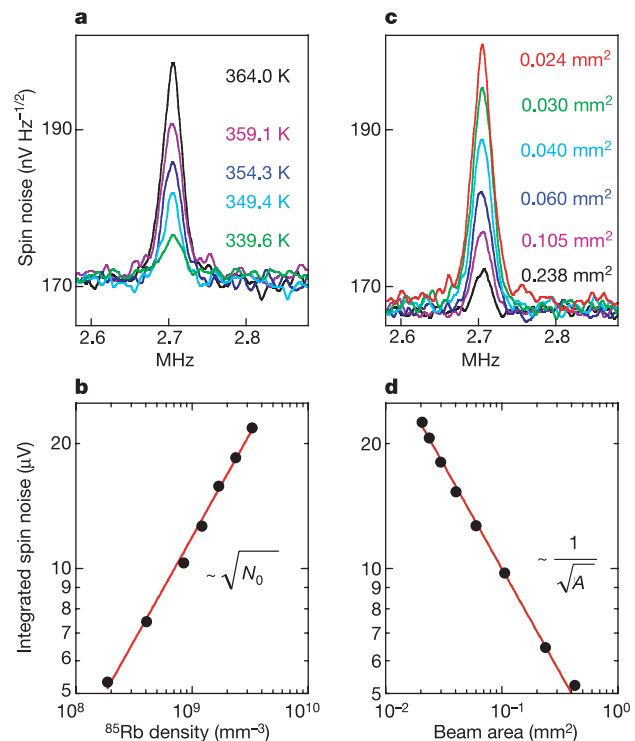


Figure 3 Spin noise from ^{85}Rb versus spin density and beam cross-section. **a**, Spin fluctuation spectrum at five different spin densities, tuned via temperature, with fixed beam area (0.030 mm^2 , or 1.2 mm^3 volume), $150 \mu\text{W}$ laser power, $\Delta_{D1} = 25$ GHz detuning, and $B = 5.8$ G. **b**, The corresponding integrated spin noise versus spin density, showing scaling with $\sqrt{N_0}$ (red line). **c**, Increasing absolute spin noise with decreasing cross-sectional beam area (that is, area within the $1/e^2$ intensity envelope of the laser). The spin density is fixed ($T = 364$ K), with constant $145 \mu\text{W}$ laser power, $\Delta_{D1} = 25$ GHz detuning, and $B = 5.8$ G. **d**, The corresponding integrated spin noise versus beam area, showing scaling with $1/\sqrt{A}$ (red line).

sensitive to ground-state spin fluctuations through the dispersive index of refraction (as distinct from techniques in which laser noise is intentionally and resonantly absorbed by the system^{25,26}).

The amplitude of fluctuations from N uncorrelated spins scales as $\sqrt{N}$, as has been observed in atomic caesium^{4,15} where, for quantum non-demolition measurements of atomic spin, such noise naturally imposes an obstacle to precision measurements. Here, $\sqrt{N}$ scaling is explicitly verified in Fig. 3a, b by temperature-tuning the Rb vapour density N_0 . A power-law fit to the integrated spin noise with slope 1/2 shows excellent agreement. Thus, the ratio of spin noise relative to the signal that one would measure in a magnetized sample ($\sim \sqrt{N}/N$) necessarily increases in systems with fewer spins. It is noteworthy, then, that the absolute noise increases when the size of the probe laser shrinks (Fig. 3c, d). At fixed density N_0 , the effective interaction volume (and hence N) is reduced by shrinking the cross-sectional area A of the laser while maintaining constant laser power. In this case, the ^{85}Rb spin noise increases. This result may be viewed as a consequence of the $A^{-1/2}$ term in equation (1), or by considering that the Faraday rotation imposed on light passing through an intentionally magnetized system is independent of beam area, so that the effective measurement sensitivity (rotation angle per unit polarized spin, θ_F/N) is larger for narrower beams. Therefore, fluctuations of order $\sqrt{N}$ spins induce correspondingly more signal. Figure 3d shows the total spin noise versus beam area, with a power-law fit to $1/\sqrt{A}$ showing excellent agreement. These

data suggest the utility of noise spectroscopy for passive probes of small systems, where the absolute amplitude of measured fluctuations actually increases when probe size is reduced, as long as measurement sensitivity increases correspondingly. In magnetometry this situation can be realized, for example, through the use of smaller Hall bar magnetometers (since the Hall voltage is independent of area, for fixed current), or as is often the case for magneto-optical measurements, through a tighter focus.

Finally, we show that fluctuation correlation spectra can reveal detailed information about complex magnetic ground states arising from, for example, nuclear magnetism and hyperfine interactions. Figure 4a shows the spin noise spectrum in ^{39}K . With increasing magnetic field, the noise peak broadens and eventually splits into four resolvable spin coherences. This is the quadratic Zeeman effect, which originates in the gradual decoupling of the electron and nuclear spin (the hyperfine interaction) by the applied magnetic field. Within each hyperfine level (see Fig. 4b), the $2F + 1$ Zeeman levels become unequally spaced, resulting in $2F$ distinct coherences between adjacent sublevels (Supplementary Figure). The energies of the Zeeman levels are given by the Breit–Rabi formula²⁴,

$$E_{F,m_F} = \frac{-\Delta_{\text{hf}}}{2(2I+1)} - g_I \mu_N B m_F \pm \frac{\Delta_{\text{hf}}}{2} \sqrt{1 + \frac{4m_F}{2I+1}x + x^2}$$

where $x = (g_I \mu_B + g_I \mu_N)B/\Delta_{\text{hf}}$, Δ_{hf} is the zero-field hyperfine splitting, m_F is the magnetic quantum number, g_I is the nuclear g -factor, μ_N is the nuclear magneton, and the $\pm$ refers to the upper and lower ($F = I \pm \frac{1}{2}$) hyperfine state. Without nuclear effects ($g_I = 0$), transitions within the two hyperfine manifolds are degenerate and Δ_{hf} can be measured via the separation between noise peaks, $\delta\Omega = 2\Omega_0^2/\Delta_{\text{hf}}$. When nuclear terms are included, transition energies within the upper and lower hyperfine manifolds are no longer exactly degenerate. This additional nuclear Zeeman splitting $\delta E_{\text{nuclear}}$ is resolved in the noise spectrum of ^{87}Rb at 38 G, where all six allowed $\Delta F = 0$, $\Delta m_F = \pm 1$ magnetic resonance transitions are observed (Fig. 4c). The nuclear moment (μ_I) and g -factor can be determined through $\delta E_{\text{nuclear}} = 2g_I \mu_N B = 2\mu_I B/I$. Lastly, spin fluctuations in thermal equilibrium also generate spontaneous spin coherence between inter-hyperfine Zeeman levels ($\Delta F = 1$, $\Delta m_F = \pm 1$), as shown in Fig. 4d for ^{39}K . At low fields, these high-frequency noise coherences split away from the ^{39}K hyperfine frequency ($\Delta_{\text{hf}} = 461.7$ MHz) with energy $\hbar\Omega \cong \pm g_F \mu_B B$ and $\pm 3g_F \mu_B B$, providing additional means of measuring Δ_{hf} from spin fluctuations alone.

We emphasize that these measurable ground-state spin coherences arise solely from random fluctuations while in thermal equilibrium, in contrast with normal methods for magnetic resonance. Nonetheless, the same detailed spectroscopic information is revealed, in accord with the fluctuation–dissipation theorem. The non-perturbative nature and inverse scaling of absolute noise with probe size appears favourable for local noise spectroscopy of small solid-state systems, where the planar geometry of many technologically-relevant structures is well-suited to high-spatial-resolution studies. For example, in a semiconductor two-dimensional electron gas with electron density 10^{11} cm^{-2} , only $\sim 1,000$ electrons are probed in a focused $1 \mu\text{m}$ laser spot. Thus, electron spin fluctuations (relative to the signal from a polarized system) are of order one part in $\sqrt{1,000}$, as compared with only one part in $\sim \sqrt{10^9}$ in these studies. Other possibilities include spectroscopy of local magnetization fluctuations in next-generation hard-disk read/write heads^{11,27}, heterostructures for ‘spintronic’²⁸ applications and quantum information processing²⁹, or proposed noise probes of phase transitions and electron correlation in quantum Hall systems³⁰, and even single atoms¹⁹. □

Methods

Polarization fluctuations $\delta\theta_F(t)$ imposed on the transmitted laser beam (due to magnetization fluctuations in the atomic vapour) are measured in a balanced photodiode

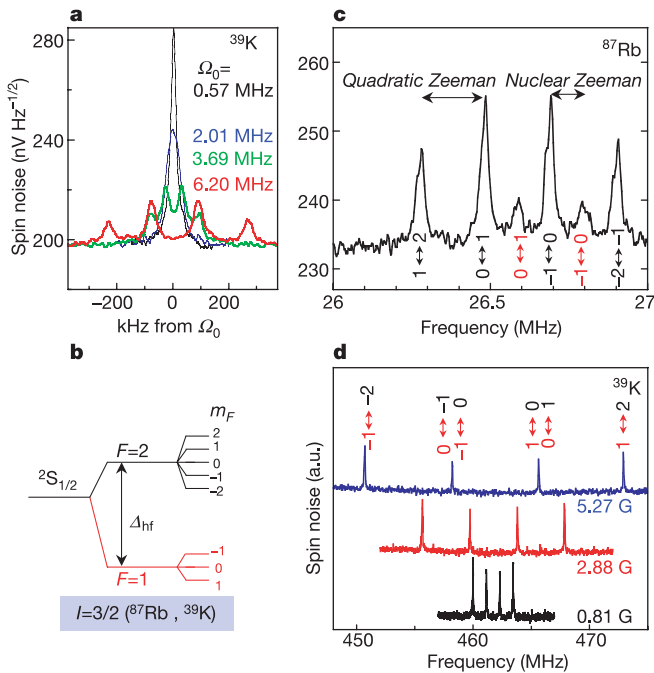


Figure 4 The ground-state Zeeman and hyperfine structure of ^{87}Rb and ^{39}K as revealed by stochastic spin coherences. **a**, Broadening and eventual breakup of the ^{39}K spin noise peak into spectrally distinct Zeeman coherences due to the quadratic Zeeman effect. The magnetic field $B = 0.81, 2.88, 5.27$ and 8.85 G, and the data are plotted relative to the centre frequency Ω_0 of the spin noise peak, as labelled. Nuclear Zeeman effects are not resolved, and the integrated spin noise is conserved (to within 5%), as expected. The laser is detuned $\Delta_{D1} = -220$ GHz from the ^{39}K D1 transition (770 nm), and $T = 456$ K. **b**, Schematic diagram of the ground-state hyperfine and Zeeman levels of ^{87}Rb ($\Delta_{\text{hf}} = 6.835$ MHz) and ^{39}K ($\Delta_{\text{hf}} = 461.7$ MHz), which both have nuclear spin $I = 3/2$. **c**, Ground-state spin fluctuation spectrum of ^{87}Rb in $B = 38.1$ G. Spin coherences between all allowed $\Delta F = 0$, $\Delta m_F = \pm 1$ Zeeman transitions are resolved, from which the hyperfine splitting and nuclear magnetic moment may be inferred. Transitions within the $F = 2$ ($F = 1$) ground-state hyperfine level are labelled in black (red). **d**, Spontaneous inter-hyperfine spin coherences ($\Delta F = 1$, $\Delta m_F = \pm 1$) in the noise spectrum of ^{39}K at $B = 0.81, 2.88$ and 5.27 G. The laser is detuned by $\Delta_{D1} = -220$ GHz, and $T = 456$ K.

bridge. The bridge consists of a polarization beamsplitter oriented at 45° to the incident laser polarization, and θ_F is measured via the normalized intensity difference between the two orthogonal components at $\pm 45^\circ$: $2\theta_F \equiv (I_{+45} - I_{-45}) / (I_{+45} + I_{-45})$. The photodiode difference current is converted to voltage (transimpedance gain = 40 V mA^{-1} , or $\sim 20 \text{ V}$ per mW of unbalanced laser power at 790 nm) and measured in a spectrum analyser. Above a few kilohertz, the measured noise floor arises primarily from photon shot noise, which contributes $\sim 175 \text{ nV Hz}^{-1/2}$ of spectrally flat (white) noise at a typical laser power of $200 \mu\text{W}$. The photodiodes and amplifier contribute an additional $65 \text{ nV Hz}^{-1/2}$ of uncorrelated white noise. All noise spectral densities are r.m.s. values, and spectra were typically signal-averaged for 10–20 min. The accuracy of measured hyperfine constants and g-factors was imposed by the $\pm 0.01 \text{ G}$ resolution of the Hall bar magnetic field sensor. Measurement of inter-hyperfine spin coherence (Fig. 4d) was performed with a higher-bandwidth, lower gain amplifier ($\sim 0.70 \text{ V mA}^{-1}$), and a typical laser power of 3.5 mW . Unless otherwise stated, measurements of rubidium (potassium) vapour were performed with 250 (125) torr of nitrogen buffer gas, which broadens the linewidth of the D1 and D2 optical transitions and causes the motion of the alkali atoms to become diffusive (thereby increasing the average time the atoms spend in the laser beam and narrowing the magnetic resonance peaks.) The diameter of the laser beam in the vapour cell can be increased (decreased) by closing (opening) the aperture in front of the final focusing lens.

Received 12 May; accepted 2 July 2004; doi:10.1038/nature02804.

- Johnson, J. B. Thermal agitation of electricity in conductors. *Nature* **119**, 50–51 (1927).
- White, D. R. *et al.* The status of Johnson noise thermometry. *Metrologia* **33**, 325–335 (1996).
- Itano, W. M. *et al.* Quantum projection noise: Population fluctuations in two-level systems. *Phys. Rev. A* **47**, 3554–3570 (1993).
- Sorensen, J. L., Hald, J. & Polzik, E. S. Quantum noise of an atomic spin polarization measurement. *Phys. Rev. Lett.* **80**, 3487–3490 (1998).
- Bloch, F. Nuclear induction. *Phys. Rev.* **70**, 460–474 (1946).
- Sleator, T., Hahn, E. L., Hilbert, C. & Clarke, J. Nuclear-spin noise. *Phys. Rev. Lett.* **55**, 1742–1745 (1985).
- Happer, W. & Mathur, B. S. Off-resonant light as a probe of optically-pumped alkali vapors. *Phys. Rev. Lett.* **18**, 577–580 (1967).
- Suter, D. & Mlynek, J. Laser excitation and detection of magnetic resonance. *Adv. Magn. Opt. Res.* **16**, 1–83 (1991).
- Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **29**, 255–284 (1966).
- Weissman, M. B. What is a spin glass? A glimpse via mesoscopic noise. *Rev. Mod. Phys.* **65**, 829–839 (1993).
- Smith, N. & Arnett, P. White-noise magnetization fluctuations in magnetoresistive heads. *Appl. Phys. Lett.* **78**, 1448–1450 (2001).
- Awschalom, D. D., DiVincenzo, D. P. & Smyth, J. F. Macroscopic quantum effects in nanometer-scale magnets. *Science* **258**, 414–421 (1992).
- Aleksandrov, E. B. & Zapsky, V. S. Magnetic resonance in the Faraday-rotation noise spectrum. *Zh. Eksp. Teor. Fiz.* **81**, 132–138 (1981).
- Mitsui, T. Spontaneous noise spectroscopy of an atomic resonance. *Phys. Rev. Lett.* **84**, 5292–5295 (2000).
- Kuzmich, A. *et al.* Quantum nondemolition measurements of collective atomic spin. *Phys. Rev. A* **60**, 2346–2350 (1999).
- Kuzmich, A., Mandel, L. & Bigelow, N. P. Generation of spin squeezing via continuous quantum nondemolition measurement. *Phys. Rev. Lett.* **85**, 1594–1597 (2000).
- Mamin, H. J., Budakian, R., Chui, B. W. & Rugar, D. Detection and manipulation of statistical polarization in small spin ensembles. *Phys. Rev. Lett.* **91**, 207604 (2003).
- Manassen, Y., Hamers, R. J., Demuth, J. E. & Castellano, A. J. Direct observation of the precession of individual paramagnetic spins on oxidized silicon surfaces. *Phys. Rev. Lett.* **62**, 2531–2534 (1989).
- Nussinov, Z., Crommie, M. F. & Balatsky, A. V. Noise spectroscopy of a single spin with spin-polarized STM. *Phys. Rev. B* **68**, 085402 (2003).
- Cleland, A. N. & Roukes, M. L. Noise processes in nanomechanical resonators. *J. Appl. Phys.* **92**, 2758–2769 (2002).
- Weaver, R. L. & Lobkis, O. I. Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies. *Phys. Rev. Lett.* **87**, 134301 (2001).
- Kastler, A. Optical methods for studying Hertzian resonances. *Science* **158**, 214–221 (1967).
- Happer, W. Optical pumping. *Rev. Mod. Phys.* **44**, 169–249 (1972).
- Corney, A. *Atomic and Laser Spectroscopy* (Clarendon, Oxford, 1977).
- Yabuzaki, T., Mitsui, T. & Tanaka, U. New type of high-resolution spectroscopy with a diode laser. *Phys. Rev. Lett.* **67**, 2453–2456 (1991).
- Ito, T., Shimomura, N. & Yabuzaki, T. Noise spectroscopy of K atoms with a diode laser. *J. Phys. Soc. Jpn* **72**, 962–963 (2003).
- Jury, J. C., Klaassen, K. B., van Peppen, J. & Wang, S. X. Measurement and analysis of noise sources in magnetoresistive sensors up to 6 GHz. *IEEE Trans. Magn.* **38**, 3545–3555 (2002).
- Wolf, S. A. *et al.* Spintronics: A spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
- Imamoglu, A. *et al.* Quantum information processing using quantum dot spins and cavity QED. *Phys. Rev. Lett.* **83**, 4204–4207 (1999).
- Joglekar, Y. N., Balatsky, A. V. & MacDonald, A. H. Noise spectroscopy and interlayer phase coherence in bilayer quantum Hall systems. *Phys. Rev. Lett.* **92**, 086803 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Littlewood, S. Gider, P. Crowell and P. Crooker for discussions. This work was supported by the Los Alamos LDRD programme.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.A.C. (crooker@lanl.gov).

Negative intrinsic resistivity of an individual domain wall in epitaxial (Ga,Mn)As microdevices

H. X. Tang¹, S. Masmanidis¹, R. K. Kawakami², D. D. Awschalom² & M. L. Roukes¹

¹Condensed Matter Physics 114-36, California Institute of Technology, Pasadena, California 91125, USA

²Department of Physics, University of California, Santa Barbara, California 93106, USA

Magnetic domains, and the boundaries that separate them (domain walls, DWs), play a central role in the science of magnetism¹. Understanding and controlling domains is important for many technological applications in spintronics, and may lead to new devices². Although theoretical efforts have elucidated several mechanisms underlying the resistance of a single DW^{3–8}, various experiments^{9–15} report conflicting results, even for the overall sign of the DW resistance. The question of whether an individual DW gives rise to an increase or decrease of the resistance therefore remains open. Here we report an approach to DW studies in a class of ferromagnetic semiconductors (as opposed to metals^{16,17}) that offer promise for spintronics¹⁸. These experiments involve microdevices patterned from monocrystalline (Ga,Mn)As epitaxial layers. The giant planar Hall effect that we previously observed¹⁹ in this material enables direct, real-time observation of the propagation of an individual magnetic DW along multiprobe devices. We apply steady and pulsed magnetic fields, to trap and carefully position an individual DW within each separate device studied. This protocol reproducibly enables high-resolution magnetoresistance measurements across an individual wall. We consistently observe negative intrinsic DW resistance that scales with channel width. This appears to originate from sizeable quantum corrections to the magnetoresistance.

Multiterminal devices for this work are patterned from $\text{Ga}_{0.948}\text{Mn}_{0.052}\text{As}$ epitaxial layers (epilayers; see Methods). The longitudinal device axis, along which the current flows, is oriented collinear with a cleave edge [110], also a cubic hard axis. Figure 1a displays the measurement set-up, showing utilization of three pairs of transverse (Hall) voltage probes to sense the magnetoresistance to the longitudinal current within the device (Methods).

(Ga,Mn)As films are intrinsically magnetized in-plane owing to a combination of compressive lattice-mismatch-induced strain and demagnetization effects²⁰. The vector plot in Fig. 1b (inset) depicts the four in-plane easy axes. A representative R – H loop (here R is the giant planar Hall resistance¹⁹, and H is magnetic field) is shown in Fig. 1b. At high negative field values the sample magnetization is saturated along easy axis I (Fig. 1b, inset). When the field is ramped up, the first jump corresponds to a magnetization transition from I to II (or [010] to [100]); the second jump completes the reversal by switching from II to III (or [100] to [010]). The magnetization transitions evolve via the formation of a 90° -DW^{19,21,22}. The square hysteresis loops obtained at low temperatures are indicative of magnetization switching dominated by wall propagation, rather than domain nucleation².

DW propagation experiments are carried out by inducing a free wall within the sample (Methods). Figure 1c shows the temporal evolution of signals obtained simultaneously using the sample's three transverse probes. The data are recorded with a drive field at point A in Fig. 1b (74 Oe). The successive jumps in the resistance versus time record for each of the three channels correspond to a single DW sequentially passing the transverse probe pairs. These

three resistance jumps display identical magnitudes and rise times, and DW propagation between adjacent pairs of voltage probes exhibits identical time intervals. Equal time delays are, in fact, expected as all probe-to-probe spacings are equal (100 μm for this measurement). These data are consistent with the picture that the DW retains a fixed shape while propagating along the device, with a speed that can be calculated directly from the time-of-flight between the probes. Detailed investigations we report elsewhere²³ demonstrate that DW velocity spanning four decades can be observed by varying the strength of the applied in-plane magnetic field. Two regimes—manifesting slow and fast dynamics, involving thermally assisted flow for low fields and viscous flow for high fields, respectively—have been identified²³.

Our initial domain wall resistance (DWR) measurements have been carried out quasistatically. First, an in-plane ‘driving’ field is adjusted so that a solitary DW propagates very slowly across the

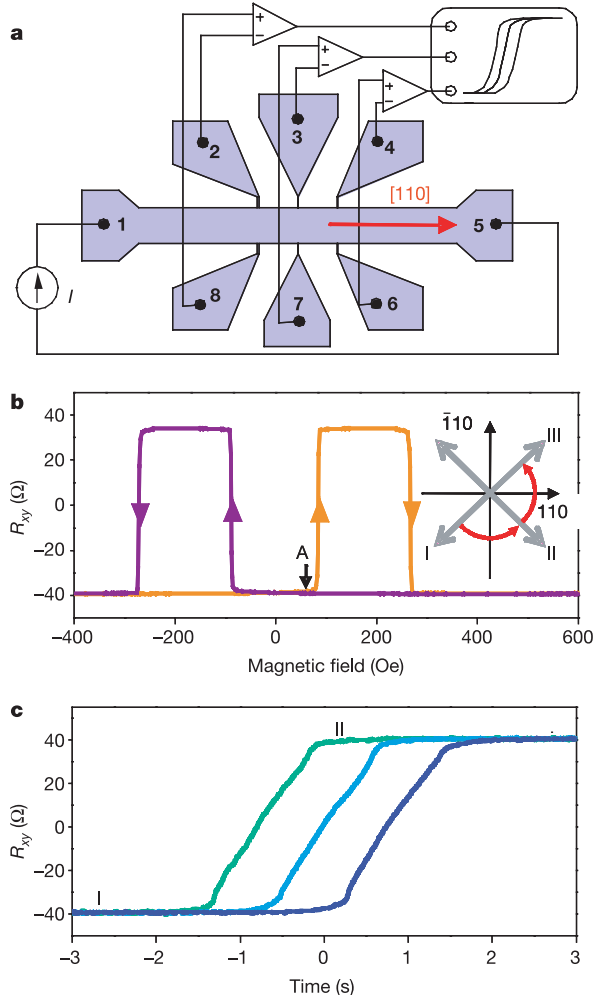


Figure 1 Sample configuration and measurement scheme. **a**, A constant sensing current is imposed between contacts 1 and 5, which are oriented along $[110]$. An external, in-plane magnetic field is applied 15° away from this axis. High-input-impedance differential amplifiers are used to make potential measurements across three transverse probes. **b**, A typical (transverse) giant planar Hall resistance, R_{xy} , versus magnetic field, H , hysteresis loop for a 100- μm -wide Hall bar at $T = 4.2$ K. Inset, schematic displaying orientation of the four in-plane easy axes of (Ga,Mn)As epilayers. **c**, After nucleation at one side of the sample, with a 74 Oe in-plane field applied, the DW propagates sequentially across the three transverse probes (differentiated by the coloured traces), successively generating transverse voltage signals.

sample (~ 50 s delay between successive probes). Then, during the DW's traverse, we obtain high-resolution measurements of two distinct longitudinal resistances, which we term R_{xx}^U and R_{xx}^D , by lock-in detection with short, 30 ms, integration times. These involve either a pair of voltage probes on the top side of the device, $R_{xx}^U = R_{15,24}$, or a pair on the bottom, $R_{xx}^D = R_{15,86}$. (Here, conventional four-probe notation is used, with $R_{ij,kl}$ corresponding to a sensing current imposed from terminal i to j , which results in an induced potential from k to l .) Simultaneously, we also measure the transverse resistances obtained from the left and right transverse probes ($R_{xy}^L = R_{15,28}$, $R_{xy}^R = R_{15,46}$) that provide continuous monitoring of the arrival and departure of individual DWs from the ‘measurement region’ between probes (Fig. 2a).

Figure 2b displays the change in longitudinal resistance resulting from entrapment of a single DW within the measurement region. A perturbation as large as 0.6% (~ 30 Ω) is manifested; however, this rather conspicuous excess resistance cannot be interpreted as the intrinsic contribution from an individual DW, because both R_{xx}^U and R_{xx}^D comprise admixtures of both the longitudinal and planar Hall resistances. A sum rule holds: the difference between R_{xx}^U and R_{xx}^D equals that between transverse resistances R_{xy}^L and R_{xy}^R (Fig. 2c). Elsewhere²⁴, we describe a theoretical model for anisotropic

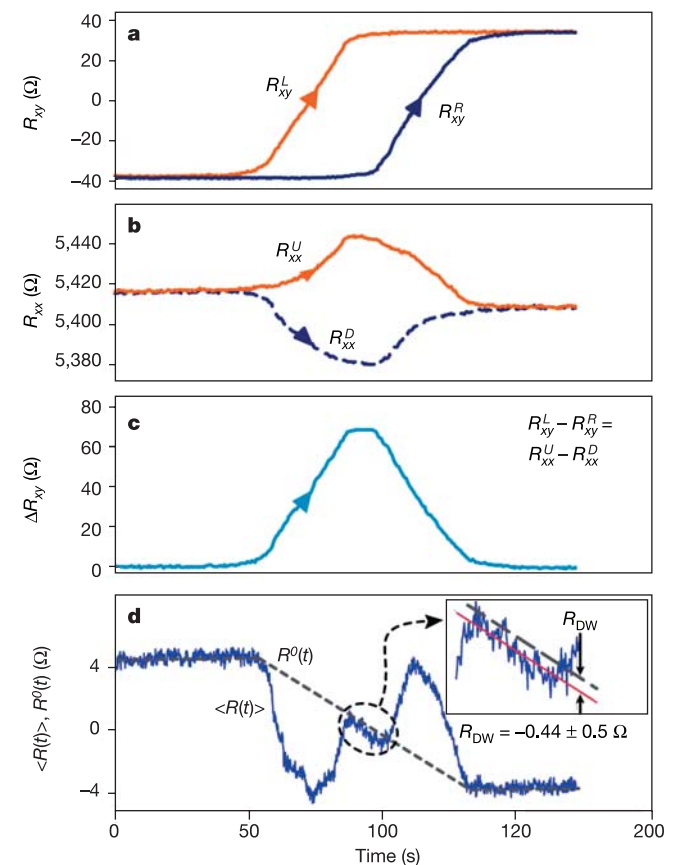


Figure 2 Time-resolved magnetoresistance measured across a single DW. **a**, Transverse resistance, R_{xy} , at 4.2 K from two pairs of probes monitoring the entrance and exit of a single DW from the channel. **b**, Longitudinal resistance, R_{xx} , across the wall is measured simultaneously from both the top and bottom of the device channel, R_{xx}^U and R_{xx}^D , respectively. **c**, A resistance sum rule is satisfied: the difference between transverse resistances, $R_{xy}^L - R_{xy}^R$, is equal to the difference between longitudinal resistances $R_{xx}^U - R_{xx}^D$. **d**, When the DW is completely resident between probes, the difference between measured longitudinal resistance, $\langle R(t) \rangle$ (violet), and a simple model described in the text, $R^0(t)$ (dashed grey) allows differentiation of the predominant, eddy-like part of domain-induced magnetoresistance from the ‘intrinsic’ DWR.

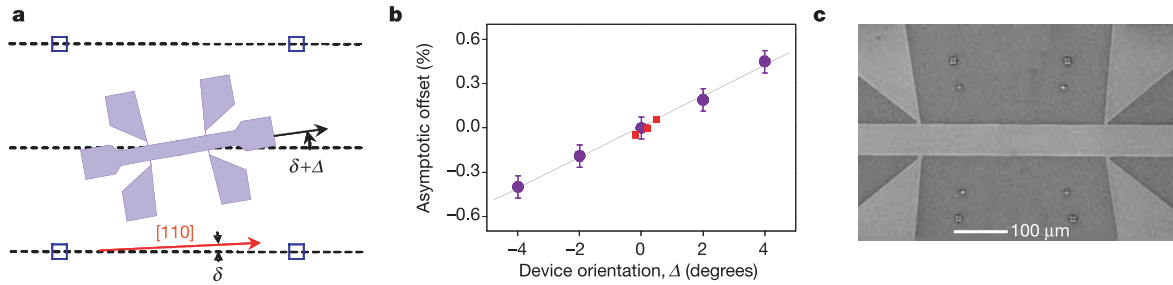


Figure 3 Scheme to align devices along the crystallographic [110] orientation. **a**, Arrays of alignment marks are patterned across a wafer with orientation along the natural ‘cleave’ axis. These are inadvertently misaligned an angle δ from the true [110] orientation. Subsequently, a series of devices are fabricated with angles stepped at 2° with respect to these alignment marks (for example, $\Delta = 0, \pm 2^\circ, \pm 4^\circ, \dots$). The zero-field asymptotic magnetoresistance between magnetization states I and II are measured for each device and the one yielding the smallest value is identified; data are shown in **b**. For the next round, we fabricate another set of devices, on the same wafer using the

same set of alignment marks, with orientation now stepped at 0.2° and centred about the optimal device identified in the first round. This process is iterated until devices with minimal offset are obtained for high-resolution investigations of DWR. Alignment accuracies of $\sim 0.03^\circ$ are realized. Data of Fig. 4 were obtained with devices oriented about $\delta \approx 0.10\text{--}0.13^\circ$, which provided the closest alignment to [110]. Error bars represent standard error of the mean. **c**, Magnified view of a $60\text{-}\mu\text{m}$ -wide device, showing the region between two transverse probes.

(‘classical’) Boltzmann transport across an individual DW. The model demonstrates that the experimentally observed longitudinal resistance perturbation can largely be attributed to the discontinuity in resistivity across a DW²⁴.

We wish to separate phenomena arising solely from the resistivity discontinuity at a DW from the more subtle magnetic scattering phenomena in that locale. Hereafter, we call the latter contributions the ‘intrinsic’ DW resistivity. Precise extraction of the intrinsic DWR can be achieved by averaging the measured longitudinal resistances, $\langle R \rangle = (R_{xx}^U + R_{xx}^D)/2$, which compensates the obfuscating contributions from the transverse voltages²⁴. In Fig. 2d we plot both $\langle R(t) \rangle$, this average of experimental data, and $R^0(t) = \rho^0(t)L/A$, a simple, idealized prediction for the longitudinal resistance. Here $R^0(t)$ and $\rho^0(t)$ describe respectively the resistance and resistivity distribution along the channel; L and A are respectively the length and cross-sectional area of the channel. For the simplest example, a device containing a single DW oriented perpendicular to the channel walls, we can write $\rho^0(x(t)) = (1 - x/L)\rho_{xx}^{(i)} + (x/L)\rho_{xx}^{(f)}$. Here we assume the DW is positioned at $0 \leq x(t) \leq L$. $\rho_{xx}^{(i)}$ and $\rho_{xx}^{(f)}$ represent the asymptotic resistivities of the initial and final magnetization states, respectively; their values differ slightly (Fig. 2b). As the DW traverses the ‘measurement region’ between longitudinal probes, $R^0(t)$ evolves linearly, reflecting the changing fractional contributions from the two domains present. The non-zero difference between the experimental trace, $\langle R(t) \rangle$, and the simple average, $R^0(t)$, clearly displays the presence of effects beyond those associated with a simple resistivity discontinuity at the DW.

There are two fundamental contributions to this difference; we call them the ‘eddy-like’ and ‘intrinsic’ DW resistivities. The former originates from local, static variations in current density induced by the resistivity discontinuity across the DW^{3,24}. (Since these contributions persist for zero DW velocity, they are distinct from true eddy currents associated with DW motion.) These eddy-like currents are only found to be appreciable when the DW is in close proximity to the probes; when the DW is roughly mid-way between these probes, $\langle R \rangle$ settles to reflect the simple linear evolution expected for $R^0(t)$ (ref. 24). However, in this ‘intermediate’ regime we observe a small, but distinct, negative offset from the expected value for $R^0(t)$. This offset evidently reflects the true, intrinsic contribution to the resistivity arising from an individual DW. In the intermediate regime, $x \approx L/2$, this particular device reproducibly manifests a negative offset $\langle R(x) \rangle - R^0(x) \approx -0.44 \Omega$. This is well within our measurement resolution ($\sim 0.2 \Omega$), but comparable to the accuracy of our interpolation yielding $R^0(t)$ (which we estimate

to be 100 p.p.m. of $\langle R \rangle$, that is, $\sim 0.5 \Omega$).

To obtain further confirmation of this negative intrinsic DW resistivity, we have refined our methodology to enhance our measurement resolution (to ~ 10 p.p.m. of the longitudinal resistance) and to minimize extraneous contributions from the bulk magnetoresistance (Methods). First, we achieve precise alignment of the devices with respect to [110] orientation to strongly suppress contributions from the anisotropic magnetoresistance (AMR). The detailed procedure is illustrated in Fig. 3. Second, instead of measuring the DW quasistatically, as in our initial measurements, we make measurements on individual, stationary DWs at zero applied field. To accomplish this we precisely position individual DWs along the device channel in a stepwise manner, using in-plane magnetic field pulses. Between these pulses we record longitudinal resistance across the stationary DW in the absence of an applied field. These zero-field measurements upon a static DW are free of effects from the field-induced longitudinal magnetoresistance, the Lorentz-force-induced component of transverse resistance, and effects from (true) eddy currents and fluctuations that can arise from DW motion.

For these high-resolution measurements, we pattern two families of long, second-generation devices (see Methods). In Fig. 4a,b we demonstrate how DW position is precisely ascertained as it is stepped along the device (see also Methods). Figure 4a displays the evolution of the transverse resistance for the left and right pair of probes from a $30\text{-}\mu\text{m}$ -wide device.

After each step in the DW’s position, we measure the longitudinal resistances R_{xx}^U and R_{xx}^D using long integration times to obtain resolution $\delta R \approx 0.1 \Omega$ ($\delta R/\langle R \rangle \approx 5 \times 10^{-6}$). Figure 4c shows the measured longitudinal resistances R_{xx}^U and R_{xx}^D . These data traces, together with their average $\langle R \rangle$ are in good agreement with the theoretical predictions²⁴. A magnified plot of $\langle R \rangle$ is displayed in Fig. 4d. For these devices, an extended phase of linear evolution is manifested in the intermediate regime, where the DW is localized between the transverse probe pairs.

Successive experiments on several devices, involving many repeated launching of individual DWs, establish firm bounds on the magnitude of this consistently negative DW resistivity. Figure 4e displays a histogram of intrinsic DWRs from 27 consecutive runs upon a precisely oriented $30\text{-}\mu\text{m}$ -wide device, taken over the course of one week. The average DWR obtained is $R_{DW} \approx -1.0 \pm 0.2 \Omega$. Two factors appear to contribute to observed variance. First, the DW orientation evidently changes slightly from run-to-run, varying between $90 \pm 20^\circ$, and therefore the length of DW changes proportionately. Second, drifts in the electronics occur over the 3-hour

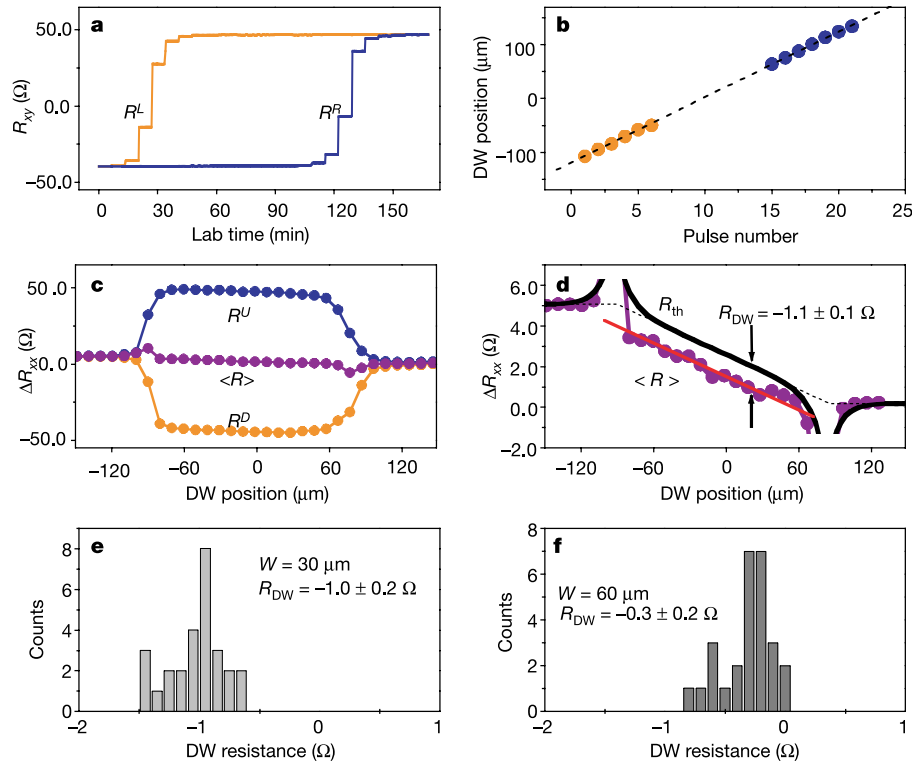


Figure 4 High-resolution extraction of the intrinsic DWR. **a**, Transverse resistance signal showing the sequential stepping of a single DW along the device channel. **b**, Calculated position of the DW as it approaches the left- and right-most pair of probes; extrapolation provides high-confidence positional data in between. **c**, Measured longitudinal resistances, R^U (blue) and R^D (orange), and their average, $\langle R \rangle$ (violet). The longitudinal axis is rescaled to reflect DW position and an average background resistance

$R_{xx} \approx 16.8 \text{ k}\Omega$ has been subtracted from all traces. **d**, Magnified view of the average resistance data (violet). The difference between the theoretical model, R^0 (black trace) and the average of data in the linear region (red line) is the DWR, $R_{DW} = -1.1 \pm 0.1 \Omega$ for this particular experiment. **e, f**, Variance of the intrinsic DWR for a large number (>25) of such experiments in which single DWs are launched and stepped along two well-aligned devices (widths, $W = 30 \mu\text{m}$ and $60 \mu\text{m}$).

measurement interval, which is required to slowly and sequentially step-and-measure each nucleated DW through a given device. Figure 4f displays measurement results on a $60\text{-}\mu\text{m}$ -wide device, for which the intrinsic contribution to the DWR is found to be $R_{DW} \approx -0.3 \pm 0.2 \Omega$. Assuming a DW width that we estimate to be $\delta_W \approx 10 \text{ nm}$ (ref. 25, $T_c = 45 \text{ K}$ is used in the calculations), these DWRs may be translated into effective DW resistivities: $\rho_{DW}(30 \mu\text{m}) = -(3 \pm 0.6) \times 10^{-4} \Omega\text{m}$ and $\rho_{DW}(60 \mu\text{m}) = -(1.8 \pm 1.2) \times 10^{-4} \Omega\text{m}$. These are quite appreciable—comparable, in fact, to the bulk resistivity of the epilayer itself, ρ . Our initial measurements (Fig. 2) yielded $\rho_{DW}(100 \mu\text{m})/\rho = -108 \pm 123\%$ for 150-nm -thick (Ga,Mn)As devices. Negative intrinsic DWR is confirmed by our subsequent high-resolution measurements on 100-nm -thick devices (Fig. 4), which yield $\rho_{DW}(30 \mu\text{m})/\rho = -100 \pm 20\%$ and $\rho_{DW}(60 \mu\text{m})/\rho = -60 \pm 40\%$.

At present, understanding of DWR is unsettled—a conflicting body of work, both experimental and theoretical, exists in the literature. DWR has, most typically, been inferred from the excess resistance arising from large ensembles of DWs within a sample^{9–11}. Progress in domain imaging and nanofabrication has more recently enabled resistance measurements on samples containing significantly reduced numbers of DWs^{12–15}. To date, however, the experimental results obtained from these various methods are conflicting; both positive^{9,11,12,14,15} and negative^{10,13} DWRs are reported. Theory has elucidated several scattering mechanisms that can yield positive DWR: reflection of carriers by the DW⁴, a ‘zigzag’ current redistribution inside the wall due to the Hall effect³, and spin-dependent scattering analogous to the GMR effect in magnetic multilayers^{5,6}. The possibility of negative DWR arising from electronic coherence

in ferromagnetic metals has also been proposed⁷. In theory, this can arise from suppression of dephasing of weakly localized electrons at a DW, thereby effectively reducing the intrinsic resistivity of DWs. Additionally, a semiclassical model, which can theoretically yield either positive or negative DWR, has been developed⁸. In this latter model, a large impurity-scattering asymmetry is required to yield a negative value.

Semiclassical models cannot account for the relatively large negative magnitude of the intrinsic DWR we observe. Although these negative DWRs could originate from quantum corrections, it is difficult to quantify the degree to which transport is phase coherent in experiments (such as these) involving ferromagnetic semiconductors. Extraction of this information from experimental data is far from straightforward; further developments in theory are clearly necessary. We anticipate that the technique for manipulating and measuring DWs reported here may be useful in ferromagnetic semiconductor devices for magnetic logic and memory¹⁶. These methods may also provide a new avenue for measuring quantum signatures of DW motion at low temperatures, where macroscopic quantum tunnelling of DWs should be manifested. □

Methods

Multiterminal device fabrication and magnetoresistance measurements

Our device fabrication begins with growth of 150-nm -thick $\text{Ga}_{0.948}\text{Mn}_{0.052}\text{As}$ epilayers on top of an insulating GaAs buffer layer by low-temperature molecular beam epitaxy (MBE). These epilayers are subsequently patterned into multiterminal devices, with longitudinal axes of the device channels (direction of current flow) oriented collinear with a cleave edge [110], which is also a cubic hard axis. Additional fabrication details are presented elsewhere^{19,26}.

Figure 1a displays the measurement set-up, showing utilization of three pairs of transverse (Hall) voltage probes to sense the magnetoresistance to a current flowing

longitudinally within the device. Measurements are carried out with a battery-supplied, constant a.c. sensing current of ≤ 500 nA; this is well below the onset of current-induced non-equilibrium effects. The induced transverse voltages are coupled through electrically isolated, low-noise differential amplifiers to a set of commonly synchronized lock-in amplifiers, which enable simultaneous acquisition of three signal channels. Both longitudinal and transverse resistances are recorded. The device is maintained at liquid helium temperature, while a magnetic field is applied within the epilayer plane via a precisely controlled, three-axis superconducting magnet. Magnetoresistance is obtained with a field oriented 15° away from $[110]$ and ramped at a rate of 15 Oe s^{-1} .

Launching of individual DWs

We first apply a strong in-plane magnetic field to saturate the magnetization, then linearly ramp to a specific field magnitude with orientation anti-aligned to the initial saturation field (for example, close to the switching field labelled as point A in Fig. 1b). At $\sim 4 \text{ K}$, DW nucleation then occurs infrequently through stochastic processes. Once nucleated, the constant in-plane field drives growth of the particular domain possessing magnetization most closely aligned with the applied field. We find that DW motion induced in this manner always involves propagation from a wide current contact pad into the channel. With this protocol, completely reproducible signals are detected.

High-resolution magnetoresistance measurements on single, stationary DWs

In the initial measurements, DWR was measured quasistatically as single DWs were sequentially driven slowly through the device. We have developed optimized techniques enabling high-resolution measurements on individual, stationary DWs. To accomplish this we have perfected a method allowing us to stepwise translate and position individual DWs along the device channel, through sequential, pulsed application of the in-plane magnetic field. Quick removal of the external field allows the DW to be 'frozen' at any desired location within the channel and, thereafter, the DW remains stationary for as long as we have been willing to measure. Subsequent application of in-plane field pulses allows the DW to be precisely stepped, in arbitrarily small increments, along the device channel. Between the applications of field pulses we are able to record longitudinal resistance across the stationary DW at zero applied field.

An individual DW is 'stepped' sequentially through the device, and its position is ascertained as follows. A magnetic field pulse—of magnitude 110 Oe , orientation 30° from hard axis $[110]$, and duration 10 s —is applied every 8 min . For a given transverse-probe pair, a jump in transverse resistance is observed after each pulse, if the DW is proximal. Thereafter, the resistance remains unchanged until the next field pulse is applied. This giant planar Hall resistance allows direct computation of the DW's relative displacement from the transverse probes. Figure 4b displays the deduced positions of a DW when it is in the vicinity of the left and right pair of probes. This linear evolution of DW position confirms that the DW travels a constant, fixed distance in response to each field pulse—about $10 \mu\text{m}$ for the aforementioned conditions. For measurement points in the intermediate region, which may be 'far' from either transverse-probe pair, DW position can be reliably determined by linear extrapolation between known positions.

To enhance our ability to resolve the intrinsic DW resistivity, we minimize the longitudinal resistance background by precise alignment of the devices with respect to $[110]$ orientation (to within $\sim 0.03^\circ$). This is achieved through a novel protocol based upon iterative, wafer-scale electron beam lithography (Fig. 3) that permits suppression of the AMR to a value as small as $\sim \Delta R/R = 6 \times 10^{-5}$. Two families of precisely aligned devices, with widths $30 \mu\text{m}$ and $60 \mu\text{m}$ and a constant $6:1$ aspect ratio (that is, length/width) from a thinner $(\text{Ga,Mn})\text{As}$ epilayer (100 nm thickness, Curie temperature $T_C \approx 45 \text{ K}$) have been iteratively patterned. The long $6:1$ aspect ratio serves to extend the intermediate region that is unperturbed by eddy-like contributions (compare Fig. 2d). To preclude spurious DW trapping along the channel itself, transverse probes are located solely at the entrance and exit of each channel (Fig. 3c).

Received 3 April; accepted 2 July 2004; doi:10.1038/nature02809.

- Hubert, A. & Schäfer, R. *Magnetic Domains: The Analysis of Magnetic Microstructures* (Springer, Berlin, 1998).
- Ferré, J. Dynamics of magnetization reversal: From continuous to patterned ferromagnetic films. *Topics Appl. Phys.* **83**, 127–165 (2002).
- Berger, L. Low-field magnetoresistance and domain drag in ferromagnets. *J. Appl. Phys.* **49**, 2156–2161 (1978).
- Cabrera, G. G. & Falicov, L. M. Theory of residual resistivity of Bloch walls. *Phys. Status Solidi B* **61**, 539–549 (1974); **62**, 217–222 (1974).
- Viret, M. *et al.* Spin scattering in ferromagnetic thin films. *Phys. Rev. B* **53**, 8464–8468 (1996).
- Levy, P. M. & Zhang, S. Resistivity due to domain wall scattering. *Phys. Rev. Lett.* **79**, 5110–5113 (1997).
- Tatara, G. & Fukumura, H. Resistivity due to a domain wall in ferromagnetic metal. *Phys. Rev. Lett.* **78**, 3773–3776 (1997).
- van Gorkom, R. P., Brataas, A. & Bauer, G. E. W. Negative domain wall resistance in ferromagnets. *Phys. Rev. Lett.* **83**, 4401–4404 (1999).
- Viret, M. *et al.* Anisotropy of domain wall resistance. *Phys. Rev. Lett.* **85**, 3962–3965 (2000).
- Ruediger, U., Yu, J., Zhang, S., Kent, A. D. & Parkin, S. S. P. Negative domain wall contribution to the resistivity of microfabricated Fe wires. *Phys. Rev. Lett.* **80**, 5639–5642 (1998).
- Klein, L. *et al.* Domain wall resistivity in SrRuO_3 . *Phys. Rev. Lett.* **84**, 6090–6093 (2000).
- Elbels, U., Radulescu, A., Henry, Y., Piroux, L. & Ounadjela, K. Spin accumulation and domain wall magnetoresistance in 35 nm Co wires. *Phys. Rev. Lett.* **84**, 983–986 (2000).
- Taniyama, T., Nakatani, L., Namikawa, T. & Yamazaki, Y. Resistivity due to domain walls in Co zigzag wires. *Phys. Rev. Lett.* **82**, 2780–2783 (1999).
- Xu, Y. B. *et al.* Magnetoresistance of a domain wall at a submicron junction. *Phys. Rev. B* **61**, R14901–R14904 (2000).
- Danneau, R. *et al.* Individual domain wall resistance in submicron ferromagnetic structures. *Phys. Rev. Lett.* **88**, 157201–157204 (2002).

- Allwood, D. A. *et al.* Submicrometer ferromagnetic NOT gate and shift register. *Science* **296**, 2003–2006 (2002).
- Ono, T. *et al.* Propagation of a magnetic domain wall in a submicrometer magnetic wire. *Science* **284**, 468–470 (1999).
- Wolf, S. A. *et al.* Spintronics: A spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
- Tang, H. X., Kawakami, R. K., Awschalom, D. D. & Roukes, M. L. Giant planar Hall effect in epitaxial $(\text{Ga,Mn})\text{As}$ devices. *Phys. Rev. Lett.* **90**, 107201–107204 (2003).
- Ohno, H. Making nonmagnetic semiconductors ferromagnetic. *Science* **281**, 951–956 (1998).
- Yamanouchi, M., Chiba, D., Matsukura, F. & Ohno, H. Current-induced domain-wall switching in a ferromagnetic semiconductor structure. *Nature* **428**, 539–542 (2004).
- Welp, U., Vlasko-Vlasov, V. K., Liu, X., Furdyna, J. K. & Wojtowicz, T. Magnetic domain structure and magnetic anisotropy in $\text{Ga}_{1-x}\text{MnAs}$. *Phys. Rev. Lett.* **90**, 167206–167209 (2003).
- Tang, H. X., Kawakami, R. K., Awschalom, D. D., Roukes, M. L. *Phys. Rev. Lett.* (submitted).
- Tang, H. X. & Roukes, M. L. Electrical transport across an individual magnetic domain wall in $(\text{Ga,Mn})\text{As}$ microdevices. *Phys. Rev. B* (submitted); preprint at (<http://arxiv.org/cond-mat/0403547>) (2004).
- Potashnik, S. J. *et al.* Saturated ferromagnetism and magnetization deficit in optimally annealed $\text{Ga}_{1-x}\text{MnAs}$ epilayers. *Phys. Rev. B* **66**, 012408 (2002).
- Tang, H. X. *Semiconductor Magnetoelectronics for Spintronics and Suspended 2DEG for Mechanoelectronics* PhD dissertation, California Inst. Technol. (2002).

Acknowledgements We acknowledge support from DARPA/DSO and AFOSR. We thank A.H. MacDonald for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.L.R. (roukes@caltech.edu).

Decline of surface temperature and salinity in the western tropical Pacific Ocean in the Holocene epoch

Lowell Stott¹, Kevin Cannariato¹, Robert Thunell², Gerald H. Haug³, Athanasios Koutavas⁴ & Steve Lund¹

¹Department of Earth Sciences, University of Southern California, Los Angeles, California 90089, USA

²Department of Geological Sciences, University of South Carolina, Columbia, South Carolina 29205, USA

³Geoforschungszentrum Potsdam, D-14473 Potsdam, Germany

⁴Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

In the present-day climate, surface water salinities are low in the western tropical Pacific Ocean and increase towards the eastern part of the basin¹. The salinity of surface waters in the tropical Pacific Ocean is thought to be controlled by a combination of atmospheric convection, precipitation, evaporation and ocean dynamics², and on interannual timescales significant variability is associated with the El Niño/Southern Oscillation cycles. However, little is known about the variability of the coupled ocean-atmosphere system on timescales of centuries to millennia. Here we combine oxygen isotope and Mg/Ca data from foraminifers retrieved from three sediment cores in the western tropical Pacific Ocean to reconstruct Holocene sea surface temperatures and salinities in the region. We find a decrease in sea surface temperatures of $\sim 0.5^\circ\text{C}$ over the past 10,000 yr, whereas sea surface salinities decreased by ~ 1.5 practical salinity units. Our data imply either that the Pacific basin as a whole has become progressively less salty or that the present salinity gradient along the Equator has developed relatively recently.

On interannual timescales, the El Niño/Southern Oscillation (ENSO) causes large changes in salinity over the equatorial Pacific as the warm, low-salinity waters from the western tropical

Pacific (WTP) are advected east into the central Pacific. This redistribution of warm waters along the Equator also alters the locus of atmospheric convection and seems to enhance the transport of heat from the tropics to higher latitudes^{2–4}. On longer timescales it is possible that other modes of variability have existed that can only be ascertained through analysis of long proxy records.

We reconstructed WTP sea surface temperature (SST) and the stable oxygen isotope composition of surface water ($\delta^{18}\text{O}_{\text{SW}}$) variability through the Holocene as a method of determining how the freshwater flux and ocean dynamics varied in the past and how that variability affected salinity gradients in the tropics. In the WTP, sea surface salinity (SSS) and $\delta^{18}\text{O}_{\text{SW}}$ are correlated, reflecting the balance between the freshwater flux (evaporation – precipitation) and salt advection by means of ocean transport. A change in the $\delta^{18}\text{O}$ of fresh water or the amount of fresh water falling on the ocean relative to inputs of salty water by means of ocean dynamics changes the salinity and also $\delta^{18}\text{O}_{\text{SW}}$. Groundwater $\delta^{18}\text{O}$ values of early, middle and late Holocene age from sites in the WTP all have the same value as present-day monsoon rainwater, indicating there has not been a change in the isotopic value of fresh water falling on the ocean over this period⁵. We therefore expect WTP $\delta^{18}\text{O}_{\text{SW}}$ values to have been controlled during the Holocene primarily by the combined effects of changing freshwater flux (the amount effect) and the flux of salty water through the WTP by means of ocean dynamics. Today these two processes result in a surface water $\delta^{18}\text{O}$ /salinity relationship that has a slope of between 0.3‰ and 0.4‰ per practical salinity unit (p.s.u.)^{6,7}.

The $\delta^{18}\text{O}_{\text{SW}}$ is incorporated into CaCO_3 shells of planktonic foraminifers that inhabit the surface ocean. To reconstruct past changes in $\delta^{18}\text{O}_{\text{SW}}$ we analysed the stable oxygen isotopic composition and Mg/Ca of calcium carbonate produced by the

surface-dwelling planktonic foraminifer *Globigerinoides ruber*. The carbonate shells of this foraminifer accumulated on the sea floor throughout the Holocene and became part of the sediment archive that was cored by the IMAGES programme in 1998 (Fig. 1). Using foraminiferal samples from three WTP IMAGES cores we reconstructed a continuous record of sea surface temperatures with Mg/Ca palaeothermometry⁸, which has an estimated uncertainty of $\pm 1^\circ\text{C}$. We also measured the $\delta^{18}\text{O}_{\text{C}}$ of the same foraminiferal material. The isotopic measurements have an analytical precision of $\pm 0.1\text{‰}$ and a practical sample reproducibility of about $\pm 0.1\text{‰}$. Combining the temperature estimates with the measured values of *G. ruber* $\delta^{18}\text{O}_{\text{C}}$, we derived values of $\delta^{18}\text{O}_{\text{SW}}$ from the carbonate palaeotemperature equation (Supplementary Information).

Our measurements are combined with previously published data from equatorial Pacific ODP Site 806 (ref. 9) to examine how surface water properties in the WTP between Indonesia and the open Pacific have varied through the Holocene epoch (Fig. 1). Samples from the Indonesian sites provide a multidecadal (MD81 and MD76) to centennial (MD70) temporal resolution through the entire Holocene, whereas the open-ocean ODP Site 806 record is of millennial resolution. Other high-quality late Pleistocene and early Holocene records are available from the Sulu Sea¹⁰ but do not provide a continuous record through the entire Holocene. The sites discussed here are all located between 6°N and 11°S (Fig. 1) and the chronology for each core is based on radiocarbon dating of fossil carbonates.

Between 10 kyr ago and the late Holocene, the $\delta^{18}\text{O}$ of *G. ruber* ($\delta^{18}\text{O}_{\text{C}}$) at each of the WTP sites decreased by $\sim 0.4\text{‰}$ (Fig. 1). Because this trend occurred after Northern Hemisphere ice sheets had mostly retreated, a change in ice volume cannot account for the

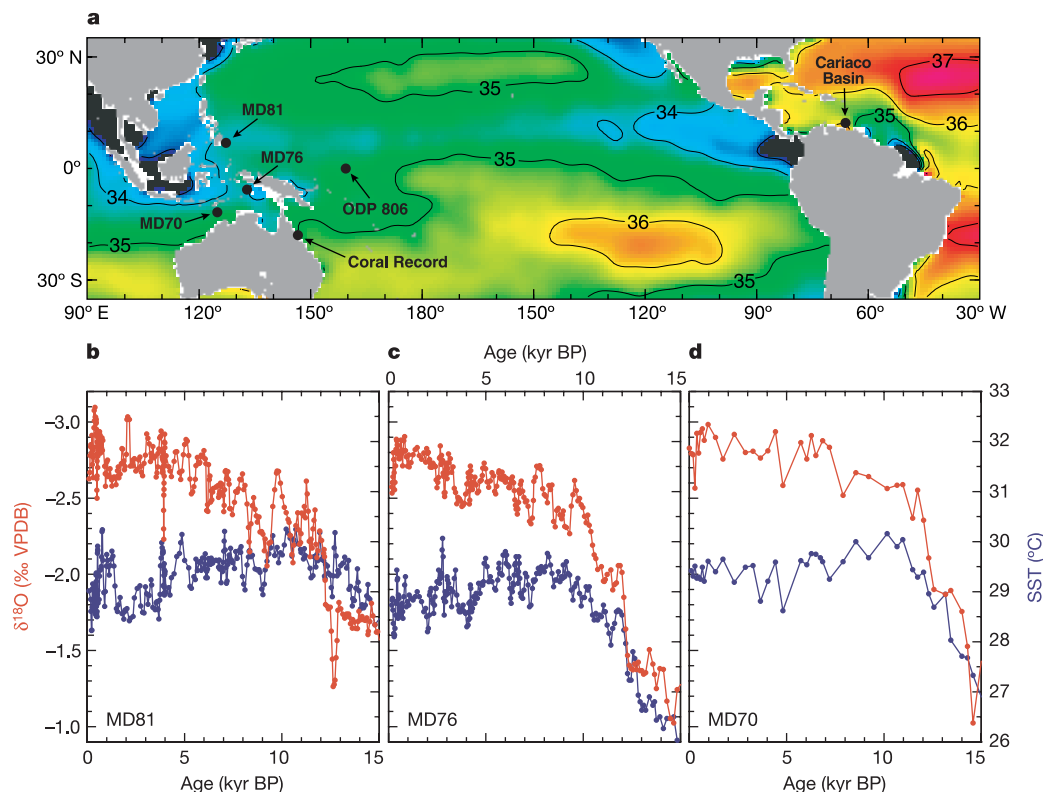


Figure 1 Western tropical Pacific marine sediment core locations and corresponding climate records. **a**, Map of mean annual surface salinities in the tropical Pacific¹ illustrating the location of IMAGES MD core locations from which the oxygen isotope (red

in **b–d**) and Mg/Ca–palaeo SST records (blue in **b–d**) were generated for this study. **b–d**, Results from individual sites: **b**, MD81; **c**, MD76; **d**, MD70. The data from sites MD81 and MD76 have been smoothed with a 3-point running mean.

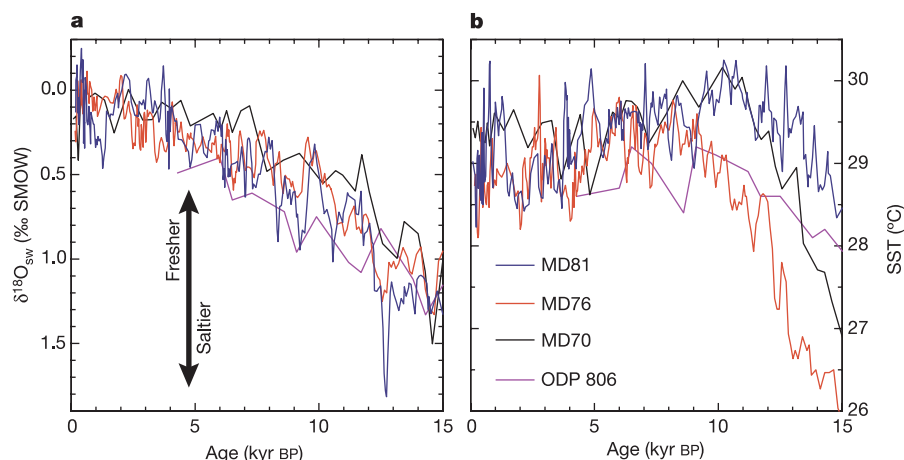


Figure 2 Reconstructed records for WTP sites. **a**, $\delta^{18}\text{O}_{\text{SW}}$; **b**, SST. The $\delta^{18}\text{O}_{\text{SW}}$ values were calculated by solving the calcite palaeotemperature equation for $\delta^{18}\text{O}_{\text{SW}}$ using the

foraminiferal Mg/Ca temperatures and the $\delta^{18}\text{O}$ values. Blue, MD81; red, MD76; black, MD70; purple, ODP 806.

large change in $\delta^{18}\text{O}_{\text{C}}$ during the late Holocene¹¹. The trend in $\delta^{18}\text{O}_{\text{C}}$ must reflect either SST warming through the Holocene, which produces a smaller isotope fractionation and lower $\delta^{18}\text{O}$ values in calcium carbonate, or a decrease in the local $^{18}\text{O}/^{16}\text{O}$ of surface waters. The Mg/Ca-derived SSTs constrain how much of the $\delta^{18}\text{O}$ trend can be attributed to temperature change. Our SST estimates indicate that surface waters across the WTP were warmest in the early Holocene ($\sim 29.5^\circ\text{C}$) and have cooled by $\sim 0.5^\circ\text{C}$ since 10 kyr ago (Fig. 2). This seems to be a robust feature of each of the records between 6°N and 11°S and is therefore unlikely to be solely a result of summer isolation changes associated with the precessional cycle. Changes in cloud albedo, evaporation and ocean dynamics arising from the precessional forcing might have contributed to the change in SST¹². But on the basis of these data alone, we cannot ascertain how these factors combined to cause the higher temperatures during the early Holocene. Nonetheless, the trend in SST does not explain the decreasing foraminiferal $\delta^{18}\text{O}_{\text{C}}$ values between the middle to late Holocene. This trend must reflect a change in the $\delta^{18}\text{O}_{\text{SW}}$ and surface salinity during the Holocene. To estimate the magnitude of the salinity change since the early Holocene we stacked the WTP $\delta^{18}\text{O}_{\text{C}}$ and Mg/Ca records, averaging the data at 250-yr intervals. The stacked data were then used to calculate $\delta^{18}\text{O}_{\text{SW}}$. Since ~ 8 kyr ago, $\delta^{18}\text{O}_{\text{SW}}$ values have decreased in the WTP by 0.5‰ (Fig. 3). On the basis of the modern $\delta^{18}\text{O}$ –salinity relationship^{6,7}, a decrease in $\delta^{18}\text{O}_{\text{SW}}$ of 0.5‰ reflects a decrease in surface salinity of between 1 and 1.5 p.s.u. This observation implies that in the early Holocene WTP surface water was as salty as surface water in the south central equatorial Pacific today (more than 35.2 p.s.u.). This observation further implies that if the salinities in the central and eastern equatorial Pacific have not changed in association with those in the WTP, the salinity gradient that exists today across the tropical Pacific would have been absent or significantly reduced in the early to middle Holocene.

Higher $\delta^{18}\text{O}_{\text{SW}}$ values have also been recognized in a middle Holocene coral record from the Great Barrier reef¹³, indicating these higher salinities extended over large portions of the WTP (Fig. 1). The higher $\delta^{18}\text{O}_{\text{SW}}$ values from the Great Barrier Reef were interpreted to reflect enhanced evaporation over that portion of the WTP during the middle Holocene. We find it difficult to reconcile the higher $\delta^{18}\text{O}_{\text{SW}}$ values that we observe in the early to middle Holocene in the northern WTP with increased evaporation because the surface salinities in this region are so strongly influenced by summer monsoon precipitation, which most climate models indi-

cate would have been enhanced in the middle Holocene by increased solar heating in boreal summer associated with the precessional forcing¹⁴. Today, the northern WTP receives in excess of 2 m of rainfall during the summer monsoon. The isotopic composition of this water is about -6.5‰ . Groundwater of early to late Holocene age from the Philippines has the same $\delta^{18}\text{O}_{\text{SMOW}}$ as modern rainwater, indicating that the source of moisture to the region has not changed during the Holocene⁵. Therefore, if the isotopic composition of rain has not changed, an isotope mass balance calculation (with no change in ocean transport) would require there to have been a decrease in rainfall of more than 2 m (the entire seasonal rainfall amount today) to shift a 25-m-thick ocean mixed layer by the observed 0.5‰ . There is no evidence for such a drastic decrease in monsoon rainfall during the Holocene. In fact, the early to middle Holocene groundwater $\delta^{18}\text{O}_{\text{SMOW}}$ values imply that there has been a persistent and strong monsoon system throughout the Holocene¹. We conclude that the increased $\delta^{18}\text{O}_{\text{SW}}$ values in the early Holocene were not due to increased evaporation but rather were the result of either increased westward advection of salty waters through the WTP, which would imply very different salinity gradients along the Equator compared with today or that the Pacific surface waters as a whole have become progressively less salty.

With only a limited number of high-resolution Holocene $\delta^{18}\text{O}_{\text{SW}}$ records available from other parts of the global ocean it is not

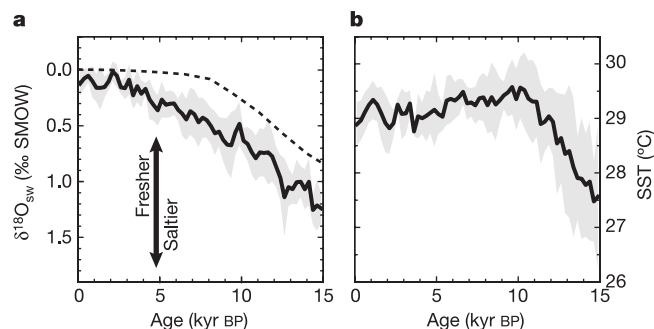


Figure 3 Stacked (average value) **(a)** $\delta^{18}\text{O}_{\text{SW}}$ and **(b)** SST records from Fig. 2, with 1σ uncertainty envelope (grey). Dashed line in panel **a** is the $\delta^{18}\text{O}_{\text{SW}}$ change due to the melting ice sheets reconstructed from stacked benthic foraminiferal $\delta^{18}\text{O}$ records¹⁹.

possible to assess whether the observed changes in the WTP reflect a larger, basin-scale change in the isotopic composition of the ocean. However, numerical model simulations have suggested that sustained shifts in the location of the Intertropical Convergence Zone (ITCZ) over the tropical Atlantic Ocean would probably affect the vapour flux to the Pacific, and change the fresh water as well as the isotopic balance over the oceans¹⁵. A persistent displacement of the ITCZ to more northerly latitudes in summer would act to trap isotopically light vapour within the Atlantic basin and decrease the vapour gain in the Pacific. The salinity and $\delta^{18}\text{O}$ changes we document from the WTP occurred in close association with the precessional cycle and with enhanced solar heating in the northern tropics during the early Holocene, which would have tended to pull the ITCZ north off its present summer latitude¹⁴. Over the course of millennia a northerly bias in the latitude of the ITCZ and reduced vapour transport between the oceans could have affected the isotopic composition of Pacific surface waters. At present there are no continuous SSS records from the tropical Atlantic and eastern Pacific that span the entire Holocene and provide the same temporal resolution that is available for the WTP. Nonetheless, previous studies have documented an early Holocene pluvial over North Africa and a stronger Indian Ocean summer monsoon in response to Earth's precessional cycle^{16,17}. Cariaco Basin sediments also contain evidence of higher rainfall in northern South America during the early Holocene, with increasingly arid conditions developing during the past 5,000 yr (ref. 18). These data all point to tropic-wide changes in the hydrological cycle that have been attributed to a more northerly position of the ITCZ during the early Holocene in response to changes in solar radiation associated with the precessional cycle. Data from other parts of the Pacific and Atlantic will now be required for an assessment of whether the changes in the hydrologic cycle affected the salinity gradient between the Pacific and Atlantic Oceans. If so, millennial to centennial scale changes in Holocene ocean thermohaline circulation would be directly affected by ocean-atmosphere processes that have occurred in the tropics. □

Received 22 January; accepted 26 July 2004; doi:10.1038/nature02903.

- Levitus, S. & Boyer, T. P. *World Ocean Atlas 1994* Vol. 3, *Salinity* (NOAA Atlas NESDIS, US Department of Commerce, Washington DC, 1994).
- Cronin, M. F. & McPhaden, M. J. Upper ocean salinity balance in the western equatorial Pacific. *J. Geophys. Res.* **103**, 27567–27587 (1998).
- Sun, D.-Z. & Trenberth, K. E. Coordinated heat removal from the equatorial Pacific during the 1986–87 El Niño. *Geophys. Res. Lett.* **25**, 2659–2662 (1998).
- McPhaden, M. J. Genesis and evolution of the 1997–98 El Niño. *Science* **283**, 950–954 (1999).
- Aggarwal, P. K., Fröhlich, K., Kulkarni, K. M. & Gourcy, L. L. Stable isotope evidence for moisture sources in the Asian summer monsoon under present and past climate regimes. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL019911 (2004).
- Fairbanks, R. G. *et al.* Evaluating climate indices and their geochemical proxies measured in corals. *Coral Reefs* **16**, S93–S100 (1997).
- Morimoto, M. *et al.* Salinity records for the 1997–98 El Niño from Western Pacific corals. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL013521 (2002).
- Nürnberg, D., Bijma, J. & Hemleben, C. Assessing the reliability of magnesium in foraminiferal calcite as a proxy of water mass temperature. *Geochim. Cosmochim. Acta* **60**, 803–814 (1996).
- Lea, D. W., Pak, D. K. & Spero, H. J. Climate impact of late Quaternary equatorial Pacific sea surface temperature variations. *Science* **289**, 1719–1724 (2000).
- Oppo, D. W., Linsley, B. K., Rosenthal, Y., Dannenmann, S. & Beaufort, L. Orbital and suborbital climate variability in the Sulu Sea, western tropical Pacific. *Geochim. Geophys. Res.* **4**, doi:10.1029/2001GC000260 (2003).
- Fairbanks, R. A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637–642 (1989).
- Clement, A. C., Seager, R. & Cane, M. A. Orbital controls on the El Niño/Southern Oscillation and the tropical climate. *Paleoceanography* **14**, 441–456 (1999).
- Gagan, M. K. *et al.* Temperature and surface-ocean water balance of the mid-Holocene tropical western Pacific. *Science* **279**, 1014–1018 (1998).
- Liu, Z., Kutzbach, J. & Wu, L. Modeling climate shift of El Niño variability in the Holocene. *Geophys. Res. Lett.* **27**, 2265–2268 (2000).
- Schmittner, A., Appenzeller, Z. & Stocker, T. F. Enhanced Atlantic freshwater export during El Niño. *Geophys. Res. Lett.* **27**, 1163–1166 (2000).
- Prell, W. L. & Van Campo, E. Coherent response of Arabian Sea upwelling and pollen transport to late Quaternary monsoonal winds. *Nature* **323**, 526–528 (1986).
- DeMenocal, P., Ortiz, J., Guilderson, T. & Sarnthein, M. Coherent high- and low-latitude climate variability during the Holocene warm period. *Science* **288**, 2198–2202 (2000).
- Haug, G. H. *et al.* Southward migration of the Intertropical Convergence Zone through the Holocene. *Science* **293**, 1304–1308 (2001).

- Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Rincon for analytical assistance. This research was supported by the US-NSF-OCE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.S. (stott@usc.edu). The foraminiferal $\delta^{18}\text{O}$ and Mg/Ca data are available at <http://www.ngdc.noaa.gov>.

Osmium isotopic constraints on the nature of the DUPAL anomaly from Indian mid-ocean-ridge basalts

S. Escrig¹, F. Capmas¹, B. Dupré² & C. J. Allègre¹

¹Laboratoire de Géochimie et Cosmochimie (UMR 7579 CNRS), Institut de Physique du Globe de Paris, Université Denis Diderot (Paris 7), 4 place Jussieu, 75252 Paris Cedex 05, France

²Laboratoire des Mécanismes et Transfert en Géologie (UMR 5563 CNRS), Université Paul Sabatier, 38 rue des Trente-Six Ponts, 31400 Toulouse, France

The isotopic compositions of mid-ocean-ridge basalts (MORB) from the Indian Ocean have led to the identification of a large-scale isotopic anomaly relative to Pacific and Atlantic ocean MORB¹. Constraining the origin of this so-called DUPAL anomaly² may lead to a better understanding of the genesis of upper-mantle heterogeneity. Previous isotopic studies^{3–10} have proposed recycling of ancient subcontinental lithospheric mantle or sediments with oceanic crust to be responsible for the DUPAL signature. Here we report Os, Pb, Sr and Nd isotopic compositions of Indian MORB from the Central Indian ridge, the Rodriguez triple junction and the South West Indian ridge. All measured samples have higher $^{187}\text{Os}/^{188}\text{Os}$ ratios than the depleted upper-mantle value^{11,12} and Pb, Sr and Nd isotopic compositions that imply the involvement of at least two distinct enriched components in the Indian upper-mantle. Using isotopic and geodynamical arguments, we reject both subcontinental lithospheric mantle and recycled sediments with oceanic crust as the cause of the DUPAL anomaly. Instead, we argue that delamination of lower continental crust may explain the DUPAL isotopic signature of Indian MORB.

In the Pb–Sr–Nd isotopic space, the Indian MORB array differs from the Pacific and Atlantic arrays by a distinct ‘low $^{206}\text{Pb}/^{204}\text{Pb}$ ’ end-member characterized by lower $^{206}\text{Pb}/^{204}\text{Pb}$, $^{143}\text{Nd}/^{144}\text{Nd}$ and higher $^{207}\text{Pb}/^{204}\text{Pb}$, $^{208}\text{Pb}/^{204}\text{Pb}$ and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios. These compositions imply the presence in the upper Indian mantle of a component that has undergone a long time-integrated evolution with high Th/U, Rb/Sr ratios and low U/Pb, Sm/Nd ratios relative to the Atlantic–Pacific upper mantle. Previous Sr–Nd–Pb isotopic studies^{3–10} of Indian MORB have related the DUPAL signature to recycled continental lithosphere, either as old subcontinental lithosphere or as sediments associated with oceanic crust. Subcontinental lithosphere requires the presence in the upper mantle of another recycled component having HIMU-like isotopic compositions (that is, high $\mu = ^{238}\text{U}/^{204}\text{Pb}$) whereas recycled oceanic crust associated with variable amounts of sediments may produce both the HIMU

and DUPAL signatures and account for all the Indian MORB isotopic variations. Solving the origin of the DUPAL anomaly remains a major goal of chemical geodynamics, essential for understanding the genesis of mantle heterogeneity.

The Re–Os isotopic system differs from other long-lived isotopic systems because the melting process strongly fractionates Re and Os; Re behaves as a moderately incompatible element and is thus enriched in melts whereas Os is strongly compatible and remains in the residue. Owing to their high Re/Os ratios, the oceanic and continental crusts develop radiogenic $^{187}\text{Os}/^{188}\text{Os}$ ratios. The radiogenic $^{187}\text{Os}/^{188}\text{Os}$ ratios measured in oceanic island basalts (OIB) are generally thought to reflect the recycling of such crustal materials in the Earth's mantle—more specifically, oceanic crust with or without sediment^{13–17}. The Re–Os system appears to be relatively insensitive to large-scale mantle metasomatism compared to the other isotopic systems, with an elevation of the mantle's $^{187}\text{Os}/^{188}\text{Os}$ ratio limited to 0.15 (ref. 18 and references therein). Indeed, the negative correlation between the Re/Os ratio and the Os content defined by metasomatized xenoliths and metasomatic minerals implies that the more Os-rich a metasomatic agent is, the less radiogenic it becomes through ^{187}Re decay and the less it affects the mantle Os isotopic evolution¹⁸. Even metasomatized, mantle fragments that have encountered significant melt extraction—that is, old continental and oceanic lithosphere—display a relative Re depletion and preserve subchondritic Os evolution. Particularly unradiogenic Os compositions of oceanic basalts, lower than the depleted MORB mantle (DMM) value, have been interpreted to reflect the presence of such an old recycled lithosphere in their mantle sources¹⁹. Thus, the Os isotopic ratios should enable us to determine which part of the continental lithosphere—that is, subcontinental mantle or continental crust-related material—is the origin of the DUPAL anomaly. The first Os isotopic analyses of MORB^{11,13,20} revealed significantly higher $^{187}\text{Os}/^{188}\text{Os}$ ratios than the estimated DMM composition^{11,12}. This difference has been attributed either to the melting of a heterogeneous mantle source^{11,20} or to sea-water-related contamination^{11,13}. Here we present evidence that, for samples with Os > 2 p.p.t., the Os isotopic compositions we measured in Indian MORB (that is, 13 samples from the Central Indian ridge (CIR), two samples from the

Rodriguez triple junction (RTJ) and one from the 39–41° E segment of the South West Indian ridge (SWIR), reported in Table 1) reflect the Indian upper-mantle composition. The ultralow Os contents of our analytical blanks allow us to interpret the measured ratios as representative of the samples we analysed. Moreover, because these Os isotopic compositions do not correlate with the Os (Fig. 1) or with MgO content, contamination by sea water ($^{187}\text{Os}/^{188}\text{Os} \approx 1.06^{21}$) and assimilation-fractional crystallization can be rejected as the origin of the high Os isotopic ratios measured. MORB from the CIR and RTJ display very homogeneous $^{187}\text{Os}/^{188}\text{Os}$ ratios (0.1356–0.1400) along the ridge, with the exception of four samples (Fig. 2). The significantly higher $^{187}\text{Os}/^{188}\text{Os}$ ratios of two CIR samples having Os contents higher than 2 p.p.t., located south of the Marie Céleste fracture zone (MCFZ), are in agreement with their significantly higher $^{87}\text{Sr}/^{86}\text{Sr}$, $^{206}\text{Pb}/^{204}\text{Pb}$ and lower $^{143}\text{Nd}/^{144}\text{Nd}$ and can be attributed to the interaction with a Réunion-type mantle flow⁸ (Fig. 2). The two other Os radiogenic CIR samples have the lowest Os contents of this study (below 2 p.p.t.) and irreproducible $^{187}\text{Os}/^{188}\text{Os}$ ratios. Their Os isotopic compositions also contrast with their other isotopic ratios, suggesting syn- or post-eruption contamination. The sample from the extreme DUPAL segment of the SWIR has the highest $^{187}\text{Os}/^{188}\text{Os}$ ratio we measured, among the most radiogenic compositions measured in MORB, and 6 p.p.t. Os. Such an extreme composition is consistent with the extreme Pb, Sr and Nd isotopic compositions of this sample. Taken as a whole, this evidence suggests that, for our data, the Os isotopic compositions of MORB with Os > 2 p.p.t. can be interpreted as reflecting the Indian upper-mantle composition and used to constrain the nature of the DUPAL anomaly.

Our Os systematics in Indian MORB confirm the radiogenic $^{187}\text{Os}/^{188}\text{Os}$ ratio of the DUPAL component²⁰ and preclude delaminated subcontinental lithospheric mantle with its low $^{187}\text{Os}/^{188}\text{Os}$ ratio as the DUPAL component. The high $^{187}\text{Os}/^{188}\text{Os}$ ratios of the extreme Indian MORB can hardly be produced by mantle metasomatism¹⁸ and instead support the recycling of a continental-crust-related component into the Indian mantle, as proposed on the basis of the Nb/U and $^{87}\text{Sr}/^{86}\text{Sr}$ correlation defined by central Indian MORB¹⁰. While sediments

Table 1 Os, Re and MgO concentrations and isotopic compositions of MORB from CIR, RTJ and 39–41° E SWIR

	Latitude	Longitude	MgO (wt%)	Re (p.p.t.)	Os (p.p.t.)	^{188}Os (p.p.t.)	$^{187}\text{Os}/^{188}\text{Os}$	$^{187}\text{Re}/^{188}\text{Os}$	$^{87}\text{Sr}/^{86}\text{Sr}$	$^{143}\text{Nd}/^{144}\text{Nd}$	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{208}\text{Pb}/^{204}\text{Pb}$
Central Indian ridge													
MD57 D14-1	–1.71	67.97	8.05	918	4.93	0.65	0.1381 (33)	899	0.702936 (06)	0.513095 (07)	18.068	15.498	37.767
MD57 D14-4	–1.71	67.97	7.82	930	17.40	2.28	0.1400 (20)	258	0.702956 (20)	0.513117 (13)	17.983	15.450	37.772
MD57 D13-m	–1.47	67.64	7.27	1148	13.39	1.75	0.1356 (15)	414	0.702983 (07)	0.513063 (07)	18.040	15.469	37.860
MD57 D11-1	–4.17	68.33	8.10	766	1.50	0.20	0.1451 (30)	2474	0.702839 (21)	0.513106 (04)	18.170	15.475	37.765
MD57 D10'-1	–6.22	68.25	7.01	963	3.16	0.41	0.1397 (15)	1470	0.703018 ^a	0.513022 ^a	18.318 ^a	15.499 ^a	38.244 ^a
MD57 D9-6	–8.00	68.08	8.97		110.93	14.53	0.1366 (06)		0.702789 (06)	0.513131 (06)	18.166	15.495	37.876
MD57 D9-1	–8.00	68.08	8.88	860	92.71	12.14	0.1400 (06)	45	0.702771 (20)	0.513108 (08)	18.210	15.536	37.946
MD57 D7-2	–12.86	66.41	8.09	865	6.41	0.84	0.1371 (24)	652	0.702847 (21)	0.513129 (42)	18.085	15.473	37.876
MD57 D7-5	–12.86	66.41	8.12	786	5.46	0.71	0.1389 (20)	701	0.702782 (16)	0.513117 (25)	18.271	15.493	37.885
MD57 D6-6	–15.86	67.28	7.94	1508	1.61	0.21	0.1479 (18)	4539	0.702907 (13)	0.513133 (06)	17.872	15.450	37.664
				604	1.34	0.18	0.1530 (19)	2175					
MD57 D5-1	–19.48	65.81	8.37	966	14.72	1.92	0.1628 (14)	318	0.703410 (18)	0.513013 (44)	18.418	15.529	38.345
				938	6.39	0.83	0.1572 (87)	711					
MD57 D3-8	–20.15	66.83	8.30	1008	9.99	1.31	0.1388 (20)	270	0.702985 (19)	0.513045 (35)	18.152	15.514	37.971
MD57 D4-3	–20.27	67.67	8.12	946	13.70	1.79	0.1468 (14)	334	0.702940 (06)	0.513106 (14)	18.294	15.495	37.946
					8.41	1.10	0.1475 (18)						
Rodriguez triple junction													
JC 3-07/3 D1	–25.57	70.03	7.87	713	10.20	1.34	0.1357 (12)	336	0.703113 ^b	0.513048 ^b	17.488 ^b	15.455 ^b	37.500 ^b
JC 2-17/2 D1	–25.61	69.94	8.24	1014	6.67	0.88	0.1388 (12)	725	0.703050 ^b	0.513027 ^b	17.477 ^b	15.455 ^b	37.470 ^b
39–41° E SWIR													
MD34 D5	–43.89	40.65	8.08 ^c	476	6.11	0.78	0.3349 (68)	385	0.704870 ^d	0.512437 ^d	16.877 ^d	15.475 ^d	37.245 ^d

Errors shown in parentheses represent 2σ and correspond to the last two digits. Analytical blanks range from 2 to 12 pg g^{–1} for Re, from 40 to 70 fg g^{–1} for Os and 0.25 to 0.45 for $^{187}\text{Os}/^{188}\text{Os}$. All samples have been blank-corrected. $^{143}\text{Nd}/^{144}\text{Nd}$ ratios are corrected for mass-discrimination using $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$ and reported to $^{143}\text{Nd}/^{144}\text{Nd} = 0.511960$ for the AMES standard. $^{87}\text{Sr}/^{86}\text{Sr}$ ratios are corrected for mass-fractionation using $^{90}\text{Sr}/^{86}\text{Sr} = 0.1194$ and normalized to $^{87}\text{Sr}/^{86}\text{Sr} = 0.71025$ for the NIST SRM987 standard. Mass fractionation for lead was 0.1‰ per atomic mass unit (a.m.u.). Typical uncertainties (1 σ) for lead ratios are 0.012 for $^{206}\text{Pb}/^{204}\text{Pb}$, 0.014 for $^{207}\text{Pb}/^{204}\text{Pb}$ and 0.031 for $^{208}\text{Pb}/^{204}\text{Pb}$. MgO contents have been determined by ICP-AES. ^{a,b,c,d} see Supplementary Information.

subducted with oceanic crust are commonly proposed to be the continental crust material recycled into the upper mantle, we develop arguments against recycled sediments as the DUPAL component and show that delamination of lower continental crust represents a better candidate.

With the aim of explaining the central Indian MORB isotopic variations, Rehkämper and Hofmann¹⁰ have related the DUPAL signature to the contamination of the Indian upper mantle with a recycled component including a 1.5-Gyr-old oceanic crust and variable proportions of pelagic sediments. To reproduce the isotopic variations of the CIR, the RTJ and the 39–41° E SWIR segment basalts, the proportion of sediments in the recycled component has to vary as a function of the degree of upper-mantle contamination. As an example, to explain the Central Indian MORB isotopic variations, the sediment contribution in the recycled component has to increase from 0.5% to 10% as the proportion of recycled component decreases (see Fig. 4 in ref. 10), which requires special pleading. Moreover, the high proportion of recycled component required in the CIR basaltic source, 100% in some cases, would imply that the Indian

upper mantle is mainly composed of recycled material. Turning to the most extreme DUPAL compositions, from the 39–41° E SWIR segment and the lavas of Afanasy Nikitin (AFN)²², even 100% of recycled component with 10% of sediments does not fully account for the data. Furthermore, the DUPAL anomaly is geographically restricted to the Indian and South Atlantic upper mantle, although recycling of oceanic crust and sediments has not been limited to this particular area. All these features argue against recycled sediments as the DUPAL component.

The CIR, the RTJ and the 39–41° E SWIR segment basalts define isotopic trends from an end-member having high $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{187}\text{Os}/^{188}\text{Os}$ and moderate $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios, close in the Pb–Sr–Nd isotopic space to the convergence zone of the Atlantic, Pacific and Indian MORB trends²³, towards different

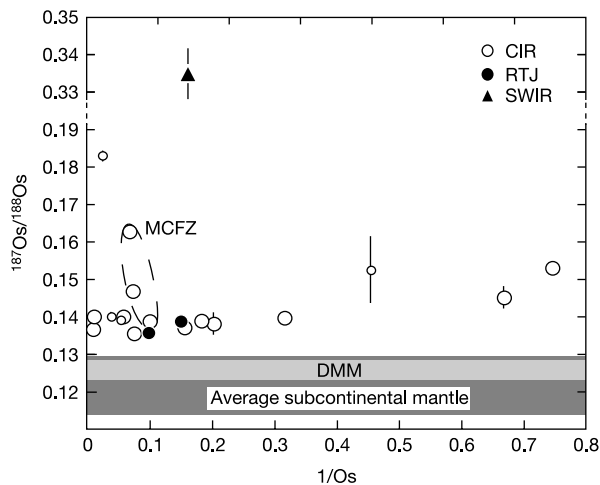


Figure 1 $1/\text{Os}$ versus $^{187}\text{Os}/^{188}\text{Os}$ ratios of MORB from CIR, RTJ and 39–41° E SWIR segment. Data from the Central Indian ridge from^{11,20} are represented as smaller symbols. The $^{187}\text{Os}/^{188}\text{Os}$ ratios of the two samples with $[\text{Os}] < 2$ p.p.t. are non-reproducible and contrast with the Pb–Sr–Nd latitudinal variation (see Fig. 2) which led us to relate these Os isotopic compositions to contamination processes. The lack of correlation between $^{187}\text{Os}/^{188}\text{Os}$ ratios and $1/\text{Os}$ or MgO content supports the absence of significant contamination of samples with $[\text{Os}] > 2$ p.p.t. by seawater interaction or assimilation-fractional crystallization. Samples located just south of the MCFZ show relatively high Os content associated with high and reproducible $^{187}\text{Os}/^{188}\text{Os}$ ratios, consistent with their Pb–Sr–Nd compositions. With the exception of these samples, MORB with $[\text{Os}] > 2$ p.p.t. from CIR and RTJ display homogeneous isotopic composition, significantly higher than the DMM value^{11,12} and higher than the average composition of subcontinental lithosphere (average $^{187}\text{Os}/^{188}\text{Os} \approx 0.1214 \pm 0.0078$ (1 σ), see the Supplementary Information for references used). Sample MD34 D5 from the 39–41° E SWIR segment presents an Os isotopic composition among the highest measured in MORB together with extreme Pb, Sr, Nd isotopic ratios. This result confirms the radiogenic Os composition of the DUPAL component. Two of our samples (MD57 D9-6 and MD57 D'10-1) have already been analysed and the compositions we measured are significantly less radiogenic than those obtained in refs 11 and 20, respectively. Such a difference may be related either to the existence of isotopic heterogeneity within MORB or to the higher Os blanks of previous Os studies. The ratios we measured in this study are consistent with the Os isotopic range of samples with $[\text{Os}] > 2$ p.p.t. and with their Pb, Sr and Nd isotopic compositions. All these features together strongly support the idea that the Os isotopic ratios for samples with $[\text{Os}] > 2$ p.p.t. reflect the MORB source composition.

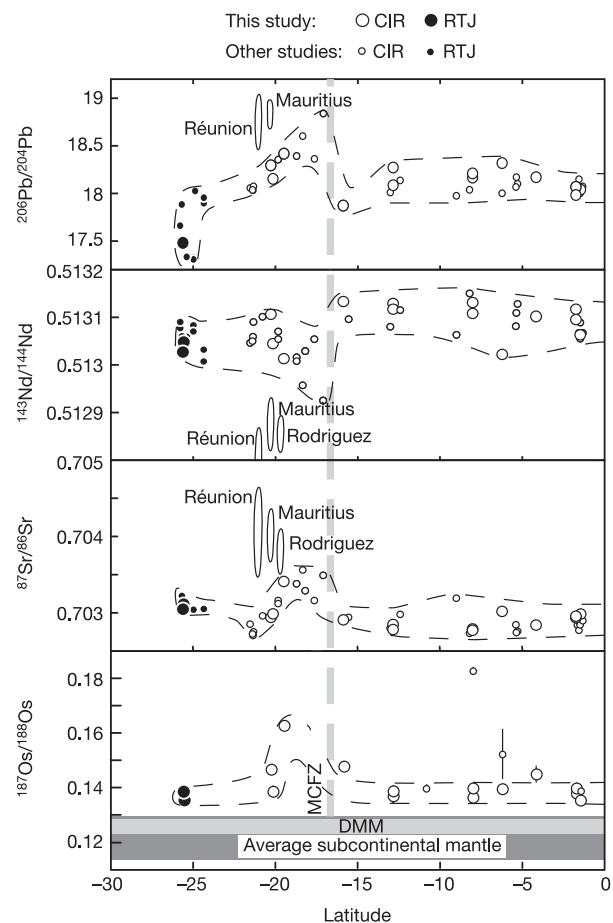


Figure 2 Latitudinal isotopic variation of CIR and RTJ basalts. The larger circles represent samples analysed for Os in this study. Smaller circles are data from previous studies (see Supplementary Information for references used). Os isotopic compositions of DMM^{11,12} and subcontinental lithosphere ($^{187}\text{Os}/^{188}\text{Os} \approx 0.1214 \pm 0.0078$ (1 σ), see Supplementary Information for references used) are also reported. Dashed curves illustrate the latitudinal isotopic variations. The ratios measured for two of our samples in previous studies and the two Os isotopic ratios of our data set thought to reflect contamination (see Fig. 1) are excluded from the Os dashed field. Os isotopic variation appears similar to those of other isotopes. Indian MORB display limited isotopic variations with latitude, except in the RTJ area and south of the MCFZ. Samples located just south of the MCFZ tend towards the isotopic fields of the Mascarene Islands (data from the GEOROC database, <http://georoc.mpch-mainz.gwdg.de/>), which supports an interaction with Mascarene-type mantle material⁶. The asymmetry of this isotopic anomaly reflects the effect of MCFZ on the mantle flow at the base of the lithosphere.

'low $^{206}\text{Pb}/^{204}\text{Pb}$ ' end-members (Fig. 3). Our data set does not provide any new constraints on the nature of the high $^{206}\text{Pb}/^{204}\text{Pb}$ component, so we focus the following discussion on the other end-members, which are responsible for the DUPAL signature. The alignment of these DUPAL end-members and the AFN (Fig. 3) supports a binary mixing between a depleted mantle source and an extreme DUPAL component.

To reconcile the continental origin of the DUPAL signature with geodynamical aspects, we favour an Indian upper-mantle contamination by delamination of lower continental crust, which represents an alternative mechanism for recycling continental material into the convecting upper mantle. The presence of lower continental crust in the MORB source has already been proposed to explain the extreme composition of one South Atlantic MORB²⁴. In a recent study²⁵, Jull and Kelemen have shown that, owing to the high temperature required for lower continental crust delamination, this delamination process is restricted to arcs, volcanic rifted margins and continental areas that are undergoing extension or removal of the underlying upper mantle. The extension associated with the Gondwana break-up has probably fulfilled the conditions required for lower continental crust delamination and such a contamination would explain the geographical distribution of the DUPAL anomaly. Lower continental crust has the advantage of long-time evolution isolated from mantle convection with high Re/

Os, Th/U, Rb/Sr and low U/Pb, Sm/Nd ratios in agreement with the isotopic features of the DUPAL end-member. Lower continental crust displays radiogenic $^{187}\text{Os}/^{188}\text{Os}$ (ref. 26) as well as low Nb/U associated with radiogenic $^{87}\text{Sr}/^{86}\text{Sr}$ (refs 27, 28), consistent with the continental signature identified in the central Indian MORB¹⁰. Mafic xenoliths sampling the African lower continental crust, located between the South Atlantic and Indian oceans, display Pb, Sr and Nd isotopic compositions^{29–31} consistent with the DUPAL signature.

Recycled lower continental crust heterogeneously distributed in the Indian upper mantle explains the different DUPAL end-member isotopic compositions. As shown in Fig. 3, each end-member could be produced by mixing in different proportions of depleted MORB-like melt and lower-continental-crust-derived melt. To maximize the contribution of the lower continental crust, the latter melt is chosen as being produced by 100% of melting of delaminated lower continental crust. Our model shows that only 10% of the melt need be derived from lower continental crust to produce the extreme compositions of RTJ basalts and 50% for the 39–41°E SWIR segment. During the melting process, mafic rocks are expected to be more fertile than the peridotitic mantle and thus to melt to a higher degree, so their contribution in the liquid is increased by a factor of three to four compared to the pristine solid mafic-ultramafic proportions³². A maximum of 4% of recycled lower

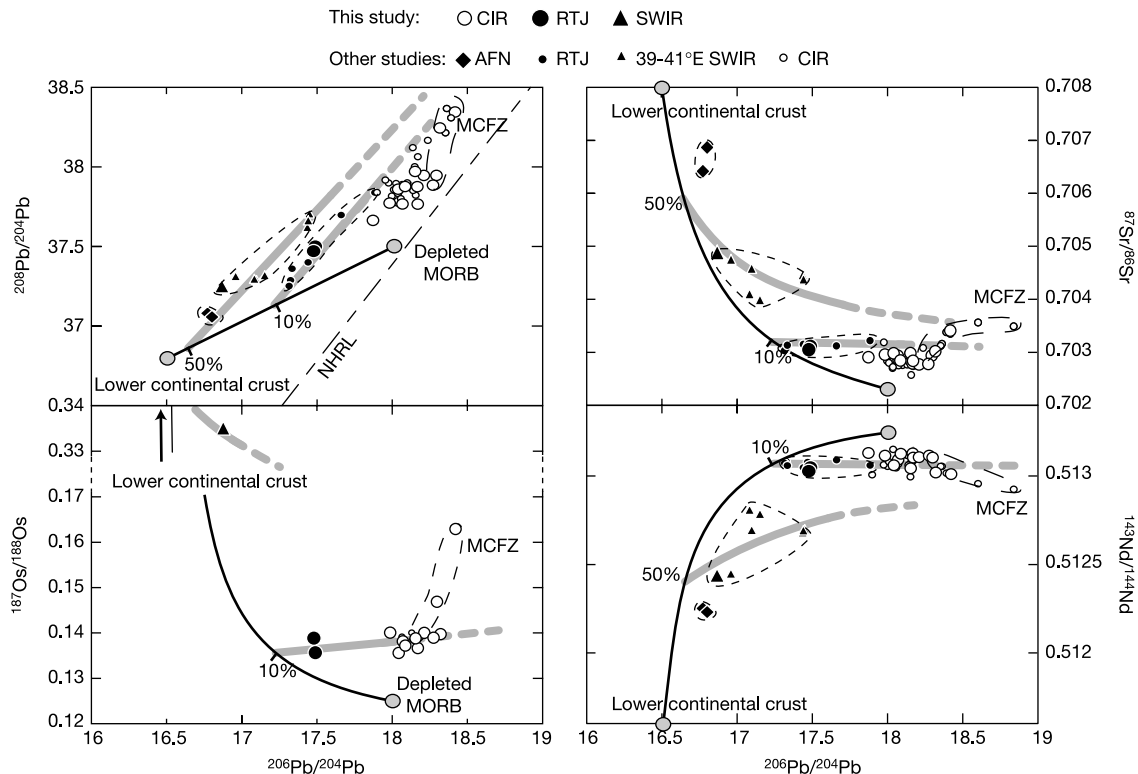


Figure 3 Pb, Sr, Nd and Os isotopic variations of the CIR, the RTJ, the 39–41°E SWIR and the AFN. NHRL, Northern Hemisphere reference line of Hart². The different groups of Indian basalts reported in the Pb–Sr–Nd isotopic diagrams define trends towards different DUPAL end-members that can be produced by mixing recycled lower continental crust and depleted upper mantle. As illustrated by the thick grey curves, the isotopic variations of basalts from the Rodriguez triple junction and the 39–41°E SWIR segment are reproduced by mixing DUPAL end-members produced by different proportions (10% and 50%) of lower continental crust with an end-member close in composition to the convergence zone of the Atlantic, Pacific and Indian MORB trends ($^{206}\text{Pb}/^{204}\text{Pb} = 19.2–19.8$; $^{208}\text{Pb}/^{204}\text{Pb} = 38.8–39.6$; $^{87}\text{Sr}/^{86}\text{Sr} = 0.703–0.704$)²³. Black curves represent

binary mixing between a MORB melt and material from the lower continental crust that is assumed to be entirely melted. Data used for mixing calculation is as follows. MORB: $^{206}\text{Pb}/^{204}\text{Pb} = 18.00$, $^{207}\text{Pb}/^{204}\text{Pb} = 15.45$, $^{208}\text{Pb}/^{204}\text{Pb} = 37.50$, $^{87}\text{Sr}/^{86}\text{Sr} = 0.7023$, $^{143}\text{Nd}/^{144}\text{Nd} = 0.51325$, $^{187}\text{Os}/^{188}\text{Os} = 0.125$, Pb = 0.45 p.p.m., Sr = 200 p.p.m., Nd = 10 p.p.m., Os = 0.15 p.p.b. Lower continental crust: $^{206}\text{Pb}/^{204}\text{Pb} = 16.50$, $^{207}\text{Pb}/^{204}\text{Pb} = 15.35$, $^{208}\text{Pb}/^{204}\text{Pb} = 36.80$, $^{87}\text{Sr}/^{86}\text{Sr} = 0.7080$, $^{143}\text{Nd}/^{144}\text{Nd} = 0.51160$, $^{187}\text{Os}/^{188}\text{Os} = 0.5$, Pb = 4.2 p.p.m.²⁸, Sr = 350 p.p.m.²⁸, Nd = 11 p.p.m.²⁸, Os = 0.04 p.p.b. The lower-continental-crust isotopic composition has been chosen within the African mafic xenoliths' isotopic range^{29–31}.

continental crust is thus sufficient to reproduce the extreme DUPAL compositions of RTJ MORB and 15% for the 39–41° E SWIR segment. Most of the MORB from the central, southeast and southwest Indian ridges has a less marked DUPAL signature that requires less than 4% of lower continental crust to be recycled into its source. The fact that recycled lower continental crust probably does not melt entirely decreases significantly the quantity of lower-continental-crust-derived melt required for a given contribution on the isotopic compositions. Indeed, incompatible elements such as Pb, Sr and Nd would be enriched in the melt derived from the lower continental crust and, because the melting degree remains high, mantle sulphides may be removed from the residue and platinum-group elements such as Os would also be enriched³³.

All of these features support the idea that delaminating of continental lithosphere, including the mafic lower continental crust, is responsible for the Indian upper-mantle isotopic anomaly. Some fragments may have sunk down to the OIB reservoir material and mixed with other recycled components, which would explain the particular compositions of Indian and South Atlantic OIB. Finally, recycling significant amounts of lower continental crust may help to explain the relatively andesitic composition of the continental crust²⁸. □

Received 3 March; accepted 30 July 2004; doi:10.1038/nature02904.

- Dupré, B. & Allègre, C. J. Pb–Sr isotope variation in Indian Ocean basalts and mixing phenomena. *Nature* **303**, 142–146 (1983).
- Hart, S. R. A large-scale isotope anomaly in the southern hemisphere mantle. *Nature* **309**, 753–757 (1984).
- Hamelin, B. & Allègre, C. J. Large-scale units in the depleted upper mantle revealed by an isotope study of the Southwest Indian Ridge. *Nature* **315**, 196–199 (1985).
- Hamelin, B., Dupré, B. & Allègre, C. J. Pb–Sr–Nd isotopic data of Indian Ocean ridges: new evidence of large-scale mapping of mantle heterogeneities. *Earth Planet. Sci. Lett.* **76**, 288–298 (1985/86).
- Michard, A., Montigny, R. & Schlich, R. Geochemistry of the mantle beneath the Rodriguez Triple Junction and the South-East Indian Ridge. *Earth Planet. Sci. Lett.* **78**, 104–114 (1986).
- Price, R. C., Kennedy, A. K., Riggs-Sneeringer, M. & Frey, F. A. Geochemistry of basalts from the Indian Ocean triple junction: implications for the generation and evolution of Indian Ocean ridge basalts. *Earth Planet. Sci. Lett.* **78**, 379–396 (1986).
- Dosso, L., Bougault, H., Beuzart, P., Calvez, J. Y. & Joron, J.-L. The geochemical structure of the South-East Indian ridge. *Earth Planet. Sci. Lett.* **88**, 47–59 (1988).
- Mahoney, J. J. *et al.* Isotopic and geochemical provinces of the Western Indian Ocean spreading centers. *J. Geophys. Res.* **94**, 4033–4052 (1989).
- Mahoney, J. J., LeRoex, A. P., Peng, Z., Fisher, R. L. & Natland, J. H. Southwestern limits of Indian Ocean ridge mantle and the origin of low ²⁰⁶Pb/²⁰⁴Pb mid-ocean ridge basalt: isotope systematics of the central Southwest Indian Ridge (17°–50°E). *J. Geophys. Res.* **97**, 19771–19790 (1992).
- Rehkämper, M. & Hofmann, A. W. Recycled ocean crust and sediment in Indian Ocean MORB. *Earth Planet. Sci. Lett.* **147**, 93–106 (1997).
- Roy-Barman, M. & Allègre, C. J. ¹⁸⁷Os/¹⁸⁶Os ratios in mid-ocean ridge basalts and abyssal peridotites. *Geochim. Cosmochim. Acta* **58**, 5043–5054 (1994).
- Snow, J. E. & Reisberg, L. Os isotopic systematics of the MORB mantle: results from altered abyssal peridotites. *Earth Planet. Sci. Lett.* **133**, 411–421 (1995).
- Martin, C. E. Osmium isotopic characteristics of mantle-derived rocks. *Geochim. Cosmochim. Acta* **55**, 1421–1434 (1991).
- Pegram, W. J. & Allègre, C. J. Osmium isotopic composition from oceanic basalts. *Earth Planet. Sci. Lett.* **111**, 59–68 (1992).
- Hauri, E. H. & Hart, S. R. Re–Os isotope systematics of HIMU and EMII oceanic island basalts from the south Pacific Ocean. *Earth Planet. Sci. Lett.* **114**, 353–371 (1993).
- Reisberg, L. *et al.* Os isotope systematics in ocean island basalts. *Earth Planet. Sci. Lett.* **120**, 149–167 (1993).
- Roy-Barman, M. & Allègre, C. J. ¹⁸⁷Os/¹⁸⁶Os in Oceanic Island Basalt: Tracing oceanic crust recycling in the mantle. *Earth Planet. Sci. Lett.* **129**, 145–161 (1995).
- Chesley, J., Righter, K. & Ruiz, J. Large-scale mantle metasomatism: a Re–Os perspective. *Earth Planet. Sci. Lett.* **219**, 49–60 (2004).
- Schaefer, B. F., Turner, S., Parkinson, I., Rogers, N. & Hawkesworth, C. Evidence for recycled Archean oceanic mantle lithosphere in the Azores plume. *Nature* **420**, 304–307 (2002).
- Schiano, P., Birck, J. L. & Allègre, C. J. Osmium–strontium–neodymium–lead isotopic covariations in mid-ocean ridge basalt glasses and heterogeneity of the upper mantle. *Earth Planet. Sci. Lett.* **150**, 363–379 (1997).
- Levasseur, S., Birck, J.-L. & Allègre, C. J. Direct measurement of femtomoles of osmium and the ¹⁸⁷Os/¹⁸⁶Os ratio in seawater. *Science* **282**, 272–274 (1998).
- Mahoney, J. J., White, W. M., Upton, B. G. J., Neal, C. R. & Scrutton, R. A. Beyond EM-1: Lavas from Afanasy-Nikitin Rise and the Crozet Archipelago, Indian Ocean. *Geology* **24**, 615–618 (1996).
- Hanan, B. B. & Graham, D. W. Lead and helium isotope evidence from oceanic basalts for a common deep source of mantle plumes. *Science* **272**, 991–995 (1996).
- Kamenetsky, V. S. *et al.* Remnants of Gondwanan continental lithosphere in oceanic upper mantle: Evidence from the South Atlantic Ridge. *Geology* **29**, 243–246 (2001).

- Jull, M. & Kelemen, P. B. On the condition for lower crust convective instability. *J. Geophys. Res.* **106**, 6423–6446 (2001).
- Saal, A. E., Rudnick, R. L., Ravizza, G. E. & Hart, S. R. Re–Os isotope evidence for the composition, formation and age of the lower continental crust. *Nature* **393**, 58–61 (1998).
- Rudnick, R. L. in *The Continental Lower Crust* (eds Fountain, D. M., Arculus, R. & Kay, R.) 269–316 (Elsevier, Amsterdam, 1992).
- Rudnick, R. L. & Fountain, D. M. Nature and composition of the continental crust: a lower crustal perspective. *Rev. Geophys.* **33**, 267–309 (1995).
- Rogers, N. W. & Hawkesworth, C. J. Proterozoic age and cumulate origin for granulite xenoliths, Lesotho. *Nature* **299**, 409–413 (1982).
- Cohen, R. S., O’Nions, R. K. & Dawson, J. B. Isotope geochemistry of xenoliths from East Africa: implications for development of mantle reservoirs and their interaction. *Earth Planet. Sci. Lett.* **68**, 209–220 (1984).
- Huang, Y.-M., Van Calsteren, P. W. & Hawkesworth, C. J. The evolution of the lithosphere in southern Africa: a perspective on the basic granulite xenoliths from kimberlites in south Africa. *Geochim. Cosmochim. Acta* **59**, 4905–4920 (1995).
- Hirschmann, M. M. & Stolper, E. M. A possible role for garnet pyroxenite in the origin of the “garnet signature” in MORB. *Contrib. Mineral. Petrol.* **124**, 185–208 (1996).
- Rehkämper, M. *et al.* Ir, Ru, Pt, and Pd in basalts and komatiites: New constraints for the geochemical behavior of the platinum-group elements in the mantle. *Geochim. Cosmochim. Acta* **63**, 3915–3934 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements S.E. thanks C.H. Langmuir for helpful discussions and its comments on the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.E. (escrig@eps.harvard.edu).

The evolution of müllerian mimicry in multispecies communities

Christopher D. Beatty, Kirsten Beirincx & Thomas N. Sherratt

Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

Prey species that are unprofitable to attack often share conspicuous colours and patterns with other coexisting defended species^{1–6}. This phenomenon, termed müllerian mimicry^{2,3}, has long been explained as a consequence of selection on defended prey to adopt a common way of advertising their unprofitability^{7,8}. However, studies using two unpalatable prey types have not always supported this theory^{9–12}. Here we show, using a system of humans hunting for computer-generated prey, that predators do not always generate strong selection for mimicry when there are two unprofitable prey types. By contrast, we demonstrate that when predators are faced with a range of different prey species, selection on unprofitable prey to resemble one another can be intense. Here the primary selective force is not one in which predators evaluate the profitabilities of distinct prey types independently, but one in which predators learn better to avoid unprofitable phenotypes that share traits distinguishing them from profitable prey^{13,14}. This need to simplify decision making readily facilitates the spread of imperfect mimetic forms from rarity, and suggests that müllerian mimicry is more likely to arise in multispecies communities.

If a predator community needs to attack a fixed number of each distinct form of defended prey (such as those with stings or toxins) before it learns to avoid them, and if this pressure is significant, then there will be selection on unprofitable prey to resemble one another^{2,3,7}. Although field experiments have lent support to the idea that common forms of unpalatable prey are at a selective

advantage over rare forms of unpalatable prey^{15–17}, the precise mechanisms involved have seldom been investigated. Experiments with avian predators feeding on two distinct forms (occurring at frequencies of 1:9 and 9:1)¹⁰, and just one form (4%, 12% and 32% of total prey)¹¹ of artificial distasteful prey have found that unpalatable phenotypes have a greater per capita probability of attack when they are rare. However, similar experiments using garden birds⁹, domestic chicks⁹ and captive great tits¹² as predators found little or no evidence to indicate that the common form is at a selective advantage over the rarer form. These findings contradict the traditional assumption that there should be selection towards uniformity because one colour pattern is easier to learn than two¹².

One reason for the above results may be that predators are not sufficiently confused to generate selection for mimicry when just two different forms are involved; they may learn rapidly to avoid both types of prey, regardless of their level of resemblance. It has recently been argued that there is likely to be far stronger selection on defended prey to adopt a common form of advertisement when there are multiple forms of prey available¹³. However, despite calls for experiments¹⁴, no study has investigated selection for mimicry when there are more than two unprofitable prey types. Natural communities often contain a wide variety of prey types that differ in their profitabilities and appearances; müllerian mimics themselves frequently participate in complex mimicry rings¹⁸. Here we used a system of humans foraging on computer-generated prey to examine the nature and intensity of selection for mimicry when there were

few forms of prey, and when there were many different types of prey available.

To evaluate the strength of selection for mimicry in simple systems containing relatively few prey phenotypes, we first conducted several related experiments (experiments 1 and 2a–c; see Methods). In experiment 1, human ‘predators’ were allowed to search a virtual environment in which they encountered individual prey items selected at random from populations of a profitable form of prey (mottled green, 40 items in total), and two unprofitable forms (mottled red and mottled blue, at nine combinations of frequencies that each totalled 40). As with previous studies^{10,11}, predators attacked more of the common form of the unprofitable prey available (Fig. 1a; number of blue attacked versus frequency of blue available $r_{38} = 0.42$, $P = 0.007$; number of red attacked versus frequency of red available $r_{38} = 0.47$, $P = 0.002$). This relationship has been explained as a consequence of predators not seeing enough of the rarest forms to complete their learning^{10,19}. However, in our study we observed how our predators behaved at each and every encounter, and noted that even rare unprofitable prey tended to be rejected by the end of the experiment (see Supplementary Table 1). This indicates that avoidance learning was largely complete, even when the unprofitable prey frequencies were relatively low. A more consistent explanation for the phenomenon may be that predators occasionally return to attacking the more common unprofitable type to assure themselves that all of that prey type were unprofitable, because there would be more to lose if some turned out to be profitable.

Despite the increase in number of unprofitable prey attacked with the frequency presented, the per capita ‘mortality’ of the unprofitable forms declined as their frequency increased (Fig. 1b). Overall, this observed predatory behaviour would be capable of generating selection for müllerian mimicry, with particularly rare forms of unprofitable prey selected to resemble common forms of unprofitable prey. Here we have elucidated selection at many more relative frequencies than have so far been examined in a single study, and can confirm, as Müller had anticipated^{2,3}, that the difference in survivorship between the rare and common forms was smaller when unprofitable prey types were similar in frequency. Unsurprisingly, significant selection for müllerian mimicry was not always evident at these intermediate frequency combinations (legend to Fig. 1b), and whether or not this was a consequence of low statistical power, it remains clear that selection under these intermediate conditions was at best weak.

In experiments 2a–c (see Fig. 2 for outline) we again investigated the strength of selection for müllerian mimicry in a simple prey community (one profitable and two unprofitable forms), this time evaluating the relative success of imperfect müllerian mimics. To do this we compared the survivorship of a distinctly coloured rare unprofitable prey (the ‘focal’ form) that was: plain, like the profitable form (experiment 2a); striped, like the more common unprofitable form (experiment 2b); or spotted so that it looked like neither the profitable nor the common unprofitable form (experiment 2c). Here, the mean per capita attack rates on each of these three forms in the three separate treatments did not differ significantly (Fig. 3, ANOVA $F_{2,27} = 2.966$, $P = 0.068$). Thus, even if selection occurred, it would not strongly favour the rare striped müllerian mimics in this simple community.

In our next experiments (experiments 3 and 4), we evaluated the strength of selection for müllerian mimicry in a more complex system with multiple forms of profitable and unprofitable prey present (Fig. 2). All of our computer-generated prey ‘species’ were distinct in appearance and were equally common (12 individuals of each). In experiments 3a–c, all of the non-focal unprofitable prey (six species) shared a common pattern element (a stripe) that was not exhibited by any of the profitable prey (six species). By contrast, in experiments 4a–c, a stripe was exhibited by three of the

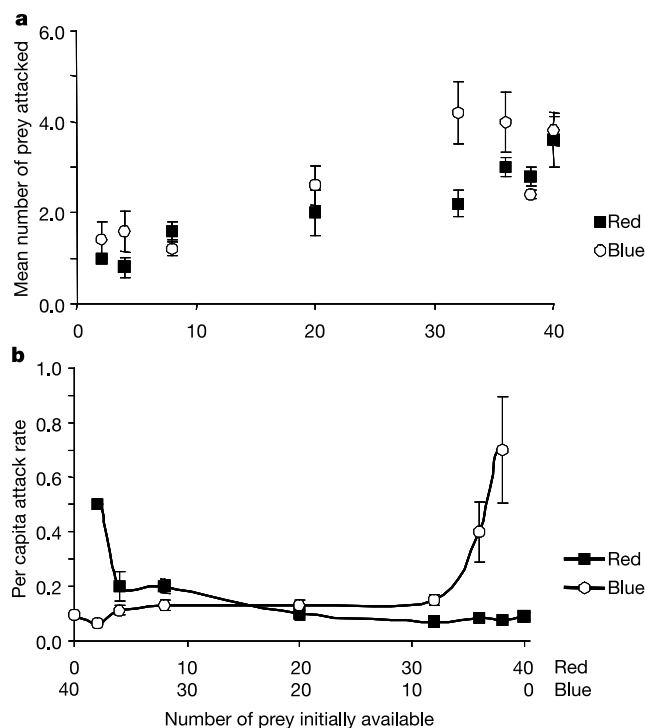


Figure 1 Results from experiment 1. **a**, The mean number of red and blue unprofitable prey attacked by predators (± 1 s.e.) in relation to the number of these prey items initially available; **b**, the implications of this behaviour for their mean per capita ‘mortality’ (± 1 s.e.). The proportion of red items attacked was significantly higher than that of blue items only when there were two red and 38 blue items (pairwise t -test on arcsin-transformed proportions $t_4 = 40.77$, $P < 0.001$). The proportion of blue items attacked was significantly higher than that of red when there were two, four or eight blue items ($t_4 = 4.58$, $P = 0.010$, $t_4 = 4.15$, $P = 0.014$, $t_4 = 7.915$, $P = 0.001$, respectively).

Experiment	Profitable prey available	Non-focal unprofitable prey available	Focal unprofitable prey form(s) available
2	 72	 72	a  [0.15] 12
			b  [0.09] 12
			c  [0.14] 12
3	 12	 12	a  [0.25] 12
			b  [0.05] 12
			c  [0.13] 12
4	 12	 12	a  [0.28] 12
			b  [0.28] 12
			c  [0.18] 12
5	 12	 12	a  [0.54] 8
			b  [0.00] 4
			c  [0.08] 8

Figure 2 A summary of experiments 2–5, in which human subjects foraged in a community of profitable and unprofitable prey. The prey forms had colours, with or without a single stripe or spot, as depicted. The numbers underneath each prey type refer to the number of that form available initially. Our analysis centred on the fate of a focal

unprofitable prey species that was magenta in colour and either plain, striped or spotted. In experiments 2–4, the focal unprofitable species was monomorphic; in experiment 5 it was dimorphic. The mean proportions attacked of each of the focal types are given in square brackets.

profitable species and three of the non-focal unprofitable species. When a stripe was reliably associated with unprofitability (experiments 3a–c), the focal prey species that carried the stripe had a far greater probability of surviving than the other forms of this species that did not (Fig. 3, ANOVA $F_{2,27} = 14.52$, $P < 0.001$; Tukey post hoc comparisons: striped versus plain, $P < 0.001$, striped versus spotted, $P = 0.018$). In these cases the focal striped prey were not only attacked far less frequently on first encounter, but they were also subsequently avoided more quickly than other forms (see Supplementary Fig. 2).

When no trait was reliably associated with unprofitability in these complex systems (experiments 4a–c), there was no significant difference in the proportions of these three different forms attacked in the separate experiments (Fig. 3; ANOVA $F_{2,27} = 1.84$, $P = 0.179$). In this case, predators continued to attack each form of the focal unprofitable prey for some time, generating low mean scores for these foragers (see Supplementary Fig. 3) and overall high prey mortality. The observation that striped prey had a greater survivorship in experiment 3b compared to 4b (ANOVA $F_{1,18} = 26.17$, $P < 0.001$) strongly suggests that predators were not simply learning to avoid all distinct unprofitable prey at equal rates. Instead, predators avoided attacking prey phenotypes at a rate related to the degree to which their traits could be categorized as unprofitable.

It has been argued that a predator's capacity to remember simultaneously the profitabilities of a wide range of prey types may be limited, and this may have an important influence on the nature of selection experienced by potential prey^{13,14}. To test this theory directly, we conducted a two-way ANOVA to compare the proportion of each focal prey type attacked in the simple (exper-

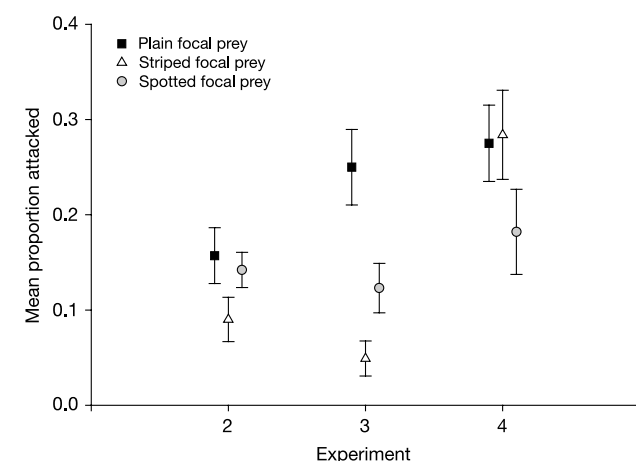


Figure 3 The mean proportions (± 1 s.e.) of each of the three forms of focal unprofitable prey attacked in experiments 2–4. An overall two-way analysis of variance on arcsin-transformed proportion of focal prey attacked revealed a highly significant interactive effect of experiment (2, 3 and 4) and focal form (plain, striped, spotted) ($F_{4,81} = 5.12$, $P = 0.001$) and highly significant main effects (experiment, $F_{2,81} = 11.91$, $P < 0.001$; focal form $F_{2,81} = 8.77$, $P < 0.001$).

iment 2) and the analogous complex (experiment 3) system. This analysis confirmed that system complexity had a significant influence on the proportion of each focal form attacked (interaction, $F_{2,54} = 3.48$, $P = 0.038$; experiment, $F_{1,54} = 0.061$, $P = 0.807$;

focal form, $F_{2,54} = 15.69$, $P < 0.001$), which seems to be driven by a combination of the increased vulnerability of the plain unprofitable form, and the decreased vulnerability of the striped (mimetic) form when the system was complex. Furthermore, given that the mimetic form was attacked at a significantly lower rate than both the plain and spotted forms when the system was complex but not when it was simple, we conclude that selection for mimicry can indeed be more intense when there are multiple species, at least under the specific conditions we have compared.

Taken together, our results suggest that there may be selection for unprofitable species in complex systems to maintain and enhance certain features (such as a colour or stripe in the right place) which happen to be shared by more unprofitable species than profitable species in a given area. To test whether an imperfect mimic could spread from rarity in a small population that contained a more common conspecific which lacked the mimetic trait, we conducted experiments 5a and 5b (Fig. 2). Despite its rarity, and the fact that it was seen at least twice by predators in all replicates, a rare form that shared a trait (stripe) with the other unprofitable species in the community had a far higher per capita survivorship than more common conspecifics that did not (experiment 5a, paired t -test on arcsin-transformed proportions $t_9 = 9.73$, $P < 0.001$). Similarly, when the striped focal form was more common than its plain conspecifics, its per capita survivorship was significantly higher (experiment 5b, $t_9 = 10.75$, $P < 0.001$). These results indicate the probable fate of intermediate phenotypes in the evolution of müllerian mimicry. If a distinctive mutant form of an unprofitable species arises, we might expect that it would suffer high mortality due to its unique appearance and extreme rarity. However, our results show that if the mutant shares simple signalling traits with more common unprofitable species, then it may well survive at higher rates than conspecifics lacking these traits.

The relative survivorship of each of our distinct focal forms differed according to the experimental context, which strongly suggests that foragers did not rely entirely on associative learning (as Müller had originally envisioned^{2,3}), but used simple rules to distinguish between profitable and unprofitable prey. Discriminative learning has been widely discussed in the psychological literature^{20–22}, but to our knowledge this is the first study to show that this type of behaviour can facilitate the spread of rare mimics with imperfect resemblance only, in a manner that is not apparent in simplified systems with relatively few prey types. As Fisher observed²³, “being recognized as unpalatable is equivalent to avoiding confusion with palatable prey”. Although Müller did not consider the appearance of profitable prey at all when making his arguments, such considerations may be essential to a full understanding of when and how müllerian mimicry evolves. □

Methods

All human ‘predators’ were visitors to the Page Break Coffee Bar situated within the MacOdrum Library at Carleton University. A total of 155 human predators (all volunteers) participated, of which 92% were non-biologists. Predators were not informed about the experimental aims and no subject was allowed to participate more than once.

The computer program was written in MS Visual Basic 6. The artificial foraging environment of each predator consisted of a grid of $n \times n$ cells. Prey (7×7 mm square and coloured/striped/spotted in a particular way according to type) were distributed within cells of this virtual grid, with no more than one prey item per cell. The predator saw only one randomly selected cell of the grid at a time (the position of the cell in the grid was not displayed) viewed in a square arena (148×148 mm) on the computer screen (see Supplementary Fig. 1). The background of the arena was comprised of a mosaic of 10% green and 90% white pixels. Predators changed cells in the search for prey by pressing a command button, causing a new cell to appear in the arena. When predators moved to a new cell that contained a prey item, they could either attack it (by clicking with the mouse cursor on it) or choose to move on without attacking it. On attacking a profitable prey item the predator gained a point and a high-pitched sound was made (which indicated profitability more effectively than using a score alone). On attacking an unprofitable prey item, the predator lost a point and a low-

pitched sound was made. Prey items that were attacked disappeared from the system. Predators were asked to forage for prey for five minutes in a way that would maximize their personal scores.

Experiment 1

Profitable prey (mottled green: 70% green, 30% white pixels, 40 items in total) and two forms of unprofitable prey (mottled red and mottled blue: 70% red and blue respectively, 30% white pixels, with a combined initial frequency of 40) were randomly distributed in a 10×10 grid. Overall, nine relative frequencies of red and blue unprofitable forms were presented (0/40, 2/38, 4/36, 8/32, 20/20, 32/8, 36/4, 38/2 and 40/0) and we allowed five different naive human predators to forage at each relative frequency (45 different human subjects in total).

Experiments 2–5

A total of 156 prey items were distributed in a 13×13 grid. Fifteen forms of prey were employed in total, consisting of blue, grey, green, red, yellow and cyan squares with and without a black stripe, and a magenta unprofitable prey type that was either plain, striped or spotted (see Fig. 2). The attack rate on the focal magenta unprofitable prey species was monitored in each of these experiments, which were all replicated ten times using different human predators (110 naive subjects in total).

Analysis

As our primary interest was in comparing the survivorship of focal prey under a given set of conditions, we used one-way analyses of variance (ANOVA) to test whether their per capita attack rates (arcsin transformed) differed according to their appearance (plain, striped, spotted). To compare the per capita attack rates (arcsin transformed) of the focal prey between experiments, we used a two-way ANOVA with experiment and focal form as fixed factors.

Using humans for foraging research

Müllerian mimicry is a taxonomically widespread phenomenon^{1–7}, so the foraging behaviours that generate it are likely to be exhibited by many different types of predator. Human subjects have long been used to test and refine ideas relating to predation^{24–26} and although humans represent visual foragers with a high capacity for learning and strategizing, they share with natural predators a finite capacity to process information^{21,27,28}. Indeed, recent work with human foragers has replicated the qualitative findings of earlier studies using great tits^{29,30}. The interpretation of our results does not depend on any behavioural traits that are unique to humans, and we hope these findings justify and inspire further work into müllerian mimicry evolution in complex communities using non-human predators.

Received 19 April; accepted 7 July 2004; doi:10.1038/nature02818.

- Bates, H. W. Contributions to an insect fauna of the Amazon valley. *Lepidoptera: Heliconidae*. *Trans. Linn. Soc. Lond.* **23**, 495–566 (1862).
- Müller, F. Über die Vortheile der Mimicry bei Schmetterlingen. *Zool. Anz.* **1**, 54–55 (1878).
- Müller, F. Ituna and Thyridia: a remarkable case of mimicry in butterflies. *Proc. Entomol. Soc. Lond.*, xx–xxix (1879).
- Sbordoni, V., Bullini, L., Scarpelli, G., Forestiero, S. & Rampini, M. Mimicry in the burnet moth *Zygaena ephialtes*: population studies and evidence of a Batesian–Müllerian situation. *Ecol. Entomol.* **4**, 83–93 (1979).
- Turner, J. R. G. In *The Biology of Butterflies* (eds Vane-Wright, R. I. & Ackery, P. R.) 141–161 (Academic, London, 1984).
- Symula, R., Schulte, R. & Summers, K. Molecular phylogenetic evidence for a mimetic radiation in Peruvian poison frogs supports a Müllerian mimicry hypothesis. *Proc. R. Soc. Lond. B* **268**, 2415–2421 (2001).
- Poulton, E. B. *The Colours of Animals: Their Meaning and Use Especially Considered in the Case of Insects* (Keegan Paul, Trench, Trubner & Co., London, 1890).
- Mallet, J. & Joron, M. Evolution of diversity in warning colour and mimicry: Polymorphisms, shifting balance, and speciation. *Annu. Rev. Ecol. Syst.* **30**, 201–233 (1999).
- Greenwood, J. J. D., Wood, E. M. & Batchelor, S. Apostatic selection of distasteful prey. *Heredity* **47**, 27–34 (1981).
- Greenwood, J. J. D., Cotton, P. A. & Wilson, D. M. Frequency-dependent selection on aposematic prey—some experiments. *Biol. J. Linn. Soc.* **36**, 213–226 (1989).
- Lindström, L., Alatalo, R. V., Lyytinen, A. & Mappes, J. Strong antiapostatic selection against novel rare aposematic prey. *Proc. Natl Acad. Sci. USA* **98**, 9181–9184 (2001).
- Rowe, C., Lindström, L. & Lyytinen, A. The importance of pattern similarity between Müllerian mimics in predator avoidance learning. *Proc. R. Soc. Lond. B* **271**, 407–413 (2004).
- MacDougall, A. & Dawkins, M. S. Predator discrimination error and the benefits of Müllerian mimicry. *Anim. Behav.* **55**, 1281–1288 (1998).
- Ruxton, G. D. Sheep in wolves’ clothing. *Nature* **394**, 833–834 (1998).
- Benson, W. W. Natural selection for Müllerian mimicry in *Heliconius erato* in Costa Rica. *Science* **176**, 936–939 (1972).
- Mallet, J. & Barton, N. H. Strong natural selection in a warning color hybrid zone. *Evolution* **43**, 421–431 (1989).
- Kapan, D. D. Three-butterfly system provides a field test of Müllerian mimicry. *Nature* **409**, 338–340 (2001).
- Mallet, J. & Gilbert, L. E. Why are there so many mimicry rings—correlations between habitat, behavior and mimicry in *Heliconius* butterflies. *Biol. J. Linn. Soc.* **55**, 159–180 (1995).
- Mallet, J. Mimicry: An interface between psychology and evolution. *Proc. Natl Acad. Sci. USA* **98**, 8928–8930 (2001).
- Shettleworth, S. J. *Cognition, Evolution, and Behavior* (Oxford Univ. Press, New York, 1998).
- Shepard, R. N., Hovland, C. I. & Jenkins, H. M. Learning and memorization of classifications. *Psychol. Monogr.* **75**(13), 1–42 (1961).

22. Pearce, J. M. in *Animal Learning and Cognition* (ed. Mackintosh, N. J.) 110–134 (Academic, San Diego, 1994).
23. Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford Univ. Press, Oxford, 1930).
24. Glanville, P. W. & Allen, J. A. Protective polymorphism in populations of computer-simulated moth-like prey. *Oikos* **80**, 565–571 (1997).
25. Götmark, F. & Hohlält, A. Bright male plumage and predation risk in passerine birds: are males easier to detect than females? *Oikos* **74**, 475–484 (1995).
26. Sherratt, T. N. & Beatty, C. D. The evolution of warning signals as reliable indicators of prey defense. *Am. Nat.* **162**, 377–389 (2003).
27. Miller, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **63**, 81–97 (1956).
28. Dukas, R. Behavioural and ecological consequences of limited attention. *Philos. Trans. R. Soc. Lond. B* **357**, 1539–1548 (2002).
29. Beatty, C. D., Bain, R. S. & Sherratt, T. N. The evolution of aggregation in profitable and unprofitable prey. *Anim. Behav.* (submitted).
30. Alatalo, R. V. & Mappes, J. Tracking the evolution of warning signals. *Nature* **382**, 708–710 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements C.D.B. and K.B. collected the data, whereas T.N.S. helped design the experiments and developed the computer program. We thank F. Batiste, A. Rashed, G. Ruxton, M. Speed, H. Van Gossum and D. Wilkinson for comments on our manuscript. The research was approved by the Carleton University Research Ethics Committee and conducted according to the guidelines set out in the Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans. The work was supported by grants to T.N.S. from NSERC, the Canada Foundation for Innovation and the Ontario Innovation Trust.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.N.S. (sherratt@ccs.carleton.ca).

Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*

Hee-Chan Seo¹, Rolf Brudvik Edvardsen¹, Anne Dorthea Maeland¹, Marianne Bjordal¹, Marit Flo Jensen¹, Anette Hansen¹, Mette Flaatt¹, Jean Weissenbach³, Hans Lehrach², Patrick Wincker³, Richard Reinhardt² & Daniel Chourrout¹

¹Sars Centre for Marine Molecular Biology, Bergen High Technology Centre, Thormøhlensgaten 55, 5008 Bergen, Norway

²Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany

³Genoscope-Centre National de Sequencage and CNRS UMR-8030, 91000 Evry, France

Tunicate embryos and larvae have small cell numbers and simple anatomical features in comparison with other chordates, including vertebrates. Although they branch near the base of chordate phylogenetic trees¹, their degree of divergence from the common chordate ancestor remains difficult to evaluate. Here we show that the tunicate *Oikopleura dioica* has a complement of nine *Hox* genes in which all central genes are lacking but a full vertebrate-like set of posterior genes is present. In contrast to all bilaterians studied so far, *Hox* genes are not clustered in the *Oikopleura* genome. Their expression occurs mostly in the tail, with some tissue preference, and a strong partition of expression domains in the nerve cord, in the notochord and in the muscle. In each tissue of the tail, the anteroposterior order of *Hox* gene expression evokes spatial collinearity, with several alterations. We propose a relationship between the *Hox* cluster breakdown, the separation of *Hox* expression domains, and a transition to a determinative mode of development.

Hox genes are involved in establishing morphological identities along the anteroposterior axis of bilaterians and cnidarians². Phylogenetic analysis suggests that the ancestor of all bilaterians had at least seven *Hox* genes—five anterior, one central and one posterior, according to the nomenclature of ref. 3—grouped in a genomic cluster where the gene order correlated with sequential expression along the anteroposterior axis. Both *Hox* gene sequences and the *Hox* cluster evolved in distinct animal lineages, with occasional cluster splits in invertebrate protostomes (*Drosophila*, *Caenorhabditis*), and gene losses^{4,5}. The significance of these discrete alterations in terms of developmental changes is a challenging enigma. Major gains of *Hox* genes coincided with the evolution of chordates, including the multiplication of entire clusters in vertebrates and an increment in the number of posterior paralogues up to six in vertebrates and cephalochordates^{6–8}. Recent sequencing in the tunicate ascidian *Ciona intestinalis*^{9,10} revealed only three posterior genes, which might equally represent a gain or a loss of genes. *C. intestinalis* also has only one central gene, probably reflecting a secondary loss of two genes, and its *Hox* cluster is split into five segments.

To gain further insight into the evolution of tunicate *Hox* complements, we identified the *Hox* genes of *O. dioica* and studied their expression and genomic organization. *O. dioica* belongs to the appendicularians, one of the three classes of tunicates. We recently showed that *O. dioica* has a very small (60–70 megabases), compact genome (one gene every 4 kilobases (kb))¹¹. Unlike ascidians, appendicularians conserve a chordate tail complex during the entire short life cycle (4 days at 20 °C in *O. dioica*). PCR cloning with degenerate primers, and whole-genome shotgun sequencing of both outbred and inbred populations (Supplementary Fig. S1), revealed nine candidate *Hox* genes. Full-length complementary DNA species were cloned for each of them, and phylogenetic analyses indicated that *Oikopleura* has three anterior *Hox* genes (*Hox1*, *Hox2* and *Hox4*) and six posterior *Hox* genes (*Hox9A*, *Hox9B*, *Hox10*, *Hox11*, *Hox12* and *Hox13*) (Fig. 1). Therefore, *O. dioica* and *C. intestinalis* share the same number of *Hox* genes but have markedly different *Hox* complements. *O. dioica* has lost the *Hox3* paralogue and all central genes, whereas *C. intestinalis* has probably lost some central genes. Either the posterior genes have been independently amplified in the *Oikopleura* lineage or a chordate ancestor already had a full complement of posterior *Hox* genes, which was subsequently reduced in the *Ciona* lineage.

We studied the expression pattern of each *O. dioica* *Hox* gene by *in situ* hybridization (see Methods) at 2.5 h after fertilization (tailbud stage), at 4 h (hatched tadpole) and at 6 h (mid-organogenesis). Here, we focus on the 4-h expression patterns (Fig. 2a), which were essentially identical to those seen at 2.5 h and were mostly concentrated in the tail. During late organogenesis, expression patterns evolve further and gradually include additional regions in the trunk/head. The tadpole tail consists of an epidermis, the 20 cells of a notochord, two rows of eight round muscle cells located dorsally and ventrally, and a nerve cord (about 60 neurons and support cells) placed on the left side of the notochord. Taken together, the expression patterns showed similarities to and important differences from those of other animals (Fig. 2b). Each tissue was the site of expression of only a subset of the nine *Hox* genes. Conversely, each *Hox* gene was expressed in only a subset of the four tail tissues, and in extreme cases in a single cell. Overall, the expression domains of distinct *Hox* genes overlapped only weakly, except in the epidermis, and most hybridizing cells expressed a single *Hox* gene. Within this partitioned expression, the sites of expression along the anteroposterior axis showed correlation with the order of the *Hox* paralogues. There were, however, several deviations from the expression collinearity, as defined by a perfect order of the anterior expression limits (with *Hox2* as the classical exception in vertebrates and in

the amphioxus). In a total of 28 two-by-two comparisons of anterior expression limits of *Hox* genes in the four tissues, the anteroposterior order of paralogues is respected in at least 24 cases.

We performed a detailed study of the *Hox* genomic organization by screening a bacterial artificial chromosome (BAC) library with nine distinct *Hox* cDNA probes. Strikingly, none of a total of 156 positive BAC clones hybridized with more than a single *Hox* probe, indicating large distances between distinct *Hox* genes. Genomic walking on about 250 kb on each side of eight *Hox* genes failed to detect linkage between any *Hox* genes, whereas three other homeo-

box genes were attained (Fig. 3a). The *Hox* genes are therefore more dispersed in the very compact genome of *O. dioica* than are the *Hox* genes of any other animal examined thus far. To describe the genomic environment of each *Hox* gene, we fully sequenced and annotated seven BAC inserts, each containing a different *Hox* gene. Each *Hox* gene was indeed isolated and surrounded by many unrelated genes, at the usual high gene density (Fig. 3b). Because recent reports have shown a partial compartmentalization of animal genomes in large chromatin domains^{12,13}, we used cDNA cloning and *in situ* hybridization to examine whether the expression of *Hox* genes and their immediate neighbours might be co-regulated.

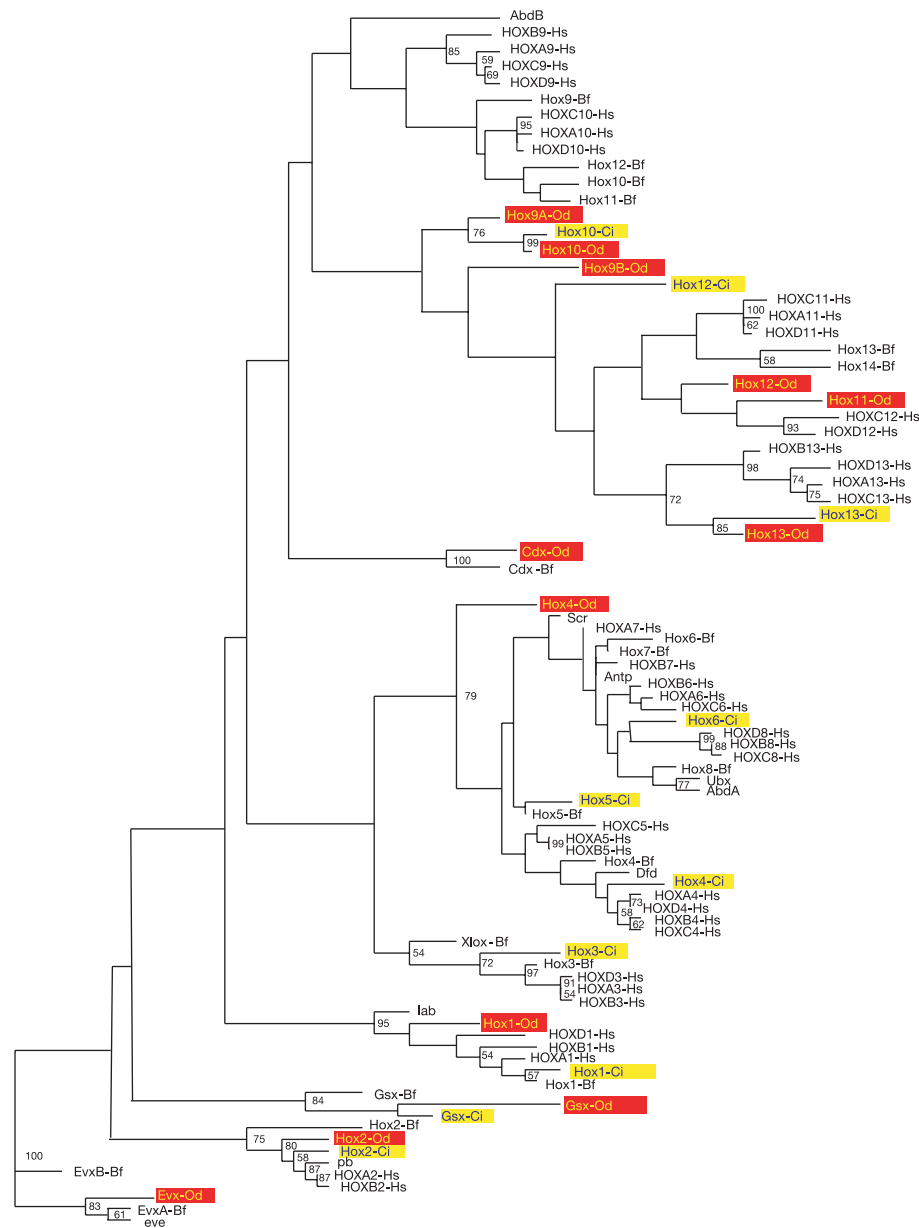


Figure 1 Evolutionary relationship of the *Hox* and *ParaHox* homeodomain sequences inferred by the neighbour-joining method. The neighbour-joining tree including a 1000 replicate bootstrap was inferred by PAUP²⁶. *O. dioica* and *C. intestinalis* gene names are highlighted in red and yellow, respectively. These include the *Okioleura ParaHox* genes *Gsx* and *Cdx* and its unique *Evx* gene, for which genomic and cDNA sequences

have been isolated. Other genes are from human (Hs), amphioxus (Bf) and *D. melanogaster*. The same data set was used for maximum-parsimony (MP; PAUP) and quartet-puzzling-likelihood (QP; TREE-PUZZLE²⁷) analyses (not shown), leading to the same overall topology. MP and QP methods placed the *O. dioica* *Hox4* gene together with the other *Hox4* sequences as a monophyletic group.

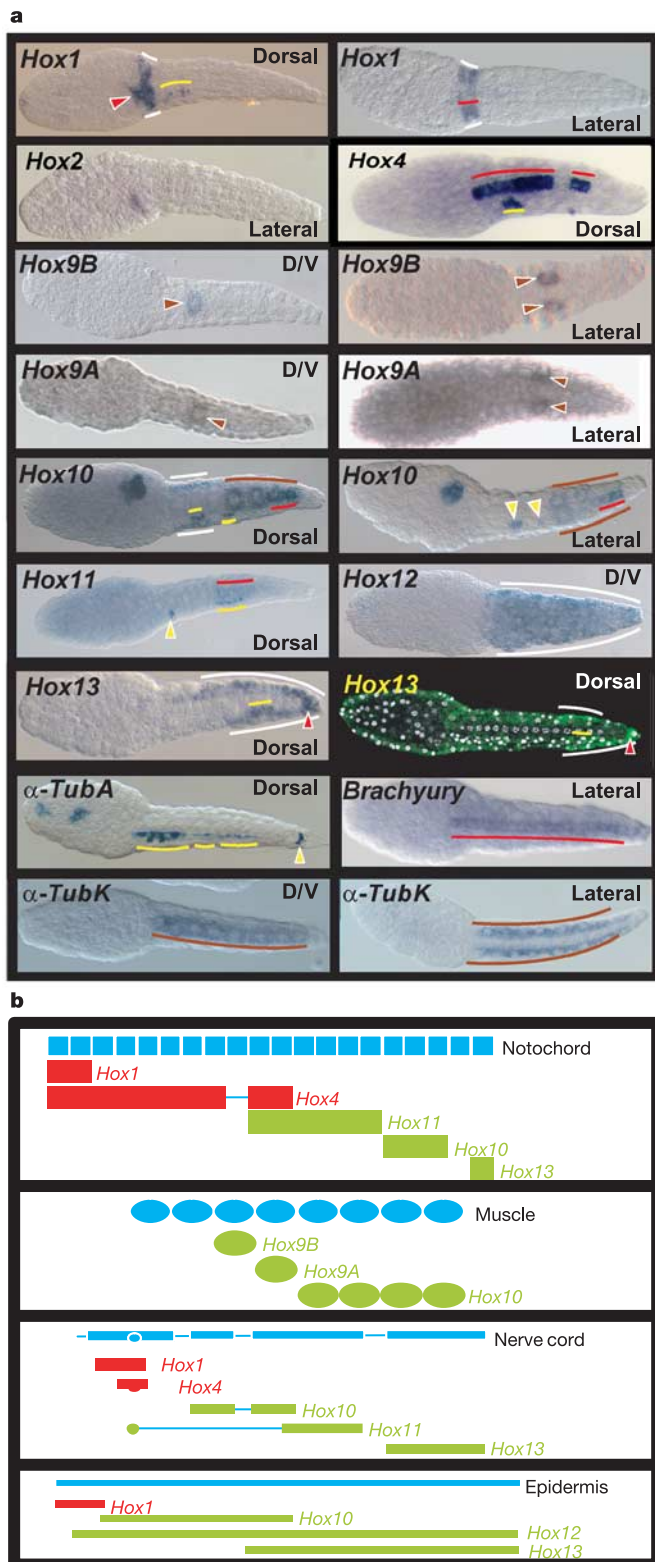


Figure 2 Expression patterns of *Oikopleura* Hox genes at 4 h after fertilization. **a**, The sites of Hox gene expression were identified by detection with both alkaline phosphatase and tyramide signal amplification (for confocal microscopy) and through comparisons with signals obtained with three marker genes (α -tubulin A (α -TubA) for neurons, α -tubulin K (α -TubK) for muscle cells, and *Brachyury* for notochord²⁶). The expression domains are identified with coloured bars or arrowheads (epidermis in white, notochord in red, nerve cord in yellow, and muscle cells in brown). **b**, The schematic organization of each tissue is drawn in blue, from the posterior end of the trunk (left) to the tail tip (right). Hox gene expression domains are represented in red for anterior genes and in green for posterior genes. The ISH protocol was adapted from ref. 25.

Almost all expression patterns of *Hox* gene neighbours were ubiquitous (not shown), ruling out local co-regulation and the possibility of an 'attraction' of *Hox* genes to specific chromatin domains. Whereas it is certain that the *Hox* cluster splits of both *C. elegans* and drosophilids were independent events, it is possible that the last common ancestor of *C. intestinalis* and *O. dioica* already had a split cluster. This question remains unanswered because we found no evidence of syntenic conservation in the *Hox* gene environments of either species. Furthermore, the expression patterns of *C. intestinalis* Hox genes are at present limited to three genes^{14–16}, two of which do not exist in *Oikopleura*, so that the conservation of expression between both tunicates could not be evaluated.

How can these results contribute to our vision of the evolution of tunicate development in chordates? First, the presence of six posterior genes in *Oikopleura* (as opposed to three in *Ciona*) indicates that posterior genes might have been amplified only once and before the radiation of chordates (Fig. 4). If this amplification is correlated with the evolution of the tail, our results are consistent with a single origin of the chordate tail, a question that has remained open until recent times^{17,18}. Second, losses of *Hox* clustering and central *Hox* genes—partial in *Ciona* and total in *Oikopleura*—are unique to tunicates among the chordates (Fig. 4). It has been proposed that the *Hox* cluster is ultimately required for the coordination of *Hox* expression timing¹⁹, and its integrity is indeed crucial for a proper expression schedule in the mouse²⁰. Moreover, the three species with a split *Hox* cluster (*Ciona*, *Drosophila*, *C. elegans*) do not display temporal collinearity, which is perhaps associated with their fast early development¹⁹. Similarly, several anterior and posterior *Hox* genes of *Oikopleura* are already expressed at 90 min after fertilization (not shown). Thus, a loss of *Hox* clustering could have facilitated a breakdown of temporal collinearity. If so, however, it is unlikely to be simple cause and effect because the loss of *Hox* clustering ranges from a single cluster split in *D. melanogaster* to a total disintegration in *Oikopleura*. Another possible cause of breakdown of the *Hox* cluster could be the loss of *Hox* gene function, which would relax the constraints for their coregulation: a conservation of function in axis patterning cannot be excluded in *Oikopleura*, considering the anteroposterior order of expression domains. Alternatively, the deviations from strict collinearity might reflect a reallocation of *Hox* genes to later functions, even though spatial regulation properties have surprisingly persisted. In any situation, it is noteworthy that *Hox* genes that lost their function in *C. elegans* and in *D. melanogaster* have remained tightly linked to at least another *Hox* gene, and the loss of *Hox* function therefore does not necessarily induce the *Hox* cluster breakdown.

Striking features of *Hox* expression domains in *Oikopleura* are their weak degree of overlap and a certain degree of tissue preference including strict tissue and cell specificity for three genes (*Hox9A*, *Hox9B* and *Hox12*). The complexity of *Hox* expression patterns supports the notion that the tail, despite its very small cell number and apparently simple anatomy, develops from fairly diverse genetic programmes that could allow the evolution of sophisticated tail functions (for example, pumping water or food, expansion of the house, swimming, or escape response). In a classical hypothesis based on comparisons of crustaceans with insects, a reduction of *Hox* expression overlaps is related to increased regional specialization²¹. However, the generalization of such a relationship encounters several obstacles, not the least of which is the vertebrate embryo axis, along which a rich *Hox* code establishes numerous positional identities within highly overlapping expression domains²². A hypothesis in favour of the loss of *Hox* genes in *C. elegans* is the change to determinative development^{23,24}, which rendered the patterning of the formed anteroposterior axis superfluous, or perhaps even undesirable. A parallel can be drawn with tunicates, because at least ascidians also have a lineage-driven mode of

development. It might be that the role of *Hox* genes in the specification of anterior–posterior position decreased, their role in the specification of tissue type increased, and a separation of expression domains occurred. The *Hox* cluster might have been

disorganized to facilitate or permit the separation of expression domains. The expression has become promoter-dependent, and a partial conservation of the promoters could explain why an apparent collinearity remains. □

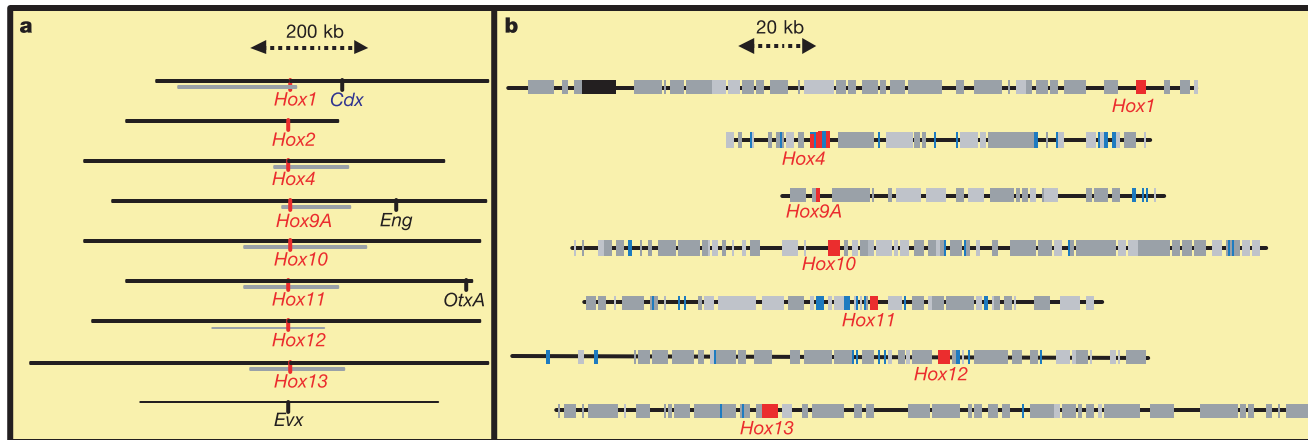


Figure 3 Genomic organization of the *Oikopleura* *Hox* genes, indicating total *Hox* cluster breakdown. **a**, A sperm BAC library (15×–20× coverage) was screened at high stringency with all nine *Hox* cDNA probes labelled with digoxigenin, and also with a cDNA probe of *Evx*. The inserts of two positive BAC clones were end-sequenced for a genomic walk (black lines). One of each *Hox*-containing BAC clone (grey lines) was fully sequenced. The *Hox2* clone could not be consistently assembled, but included no other *Hox* gene.

b, The sequence of each BAC clone was annotated using BLASTX (dark grey rectangles) and other gene prediction tools (light grey rectangles)¹¹. Each *Hox* gene was isolated in its BAC insert except *Hox4*, which was partly duplicated (see Supplementary Fig. S2). Blue bars represent short repeats conserved within or between BAC clones. A single transposable element related to *Gypsy* elements was identified in the *Hox1* BAC (black rectangle).

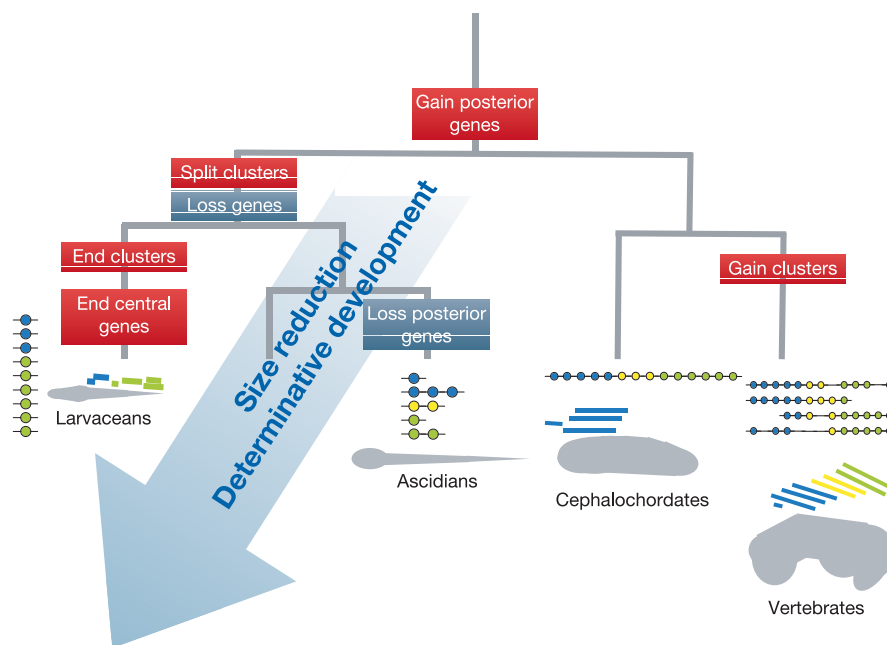


Figure 4 Discrete changes of *Hox* gene complements in chordates. The chordate ancestor gained a rich set of posterior genes, which were inherited in the three subphyla but partly lost in ascidians. Central genes were gradually lost in tunicates, with larvaceans keeping anterior and posterior genes only. Whereas the *Hox* cluster was multiplied in vertebrates (with subsequent losses of a few paralogues in some clusters), the cluster

degenerated in tunicates, and ultimately disappeared in larvaceans. The loss of central genes and of the *Hox* cluster coincides with a partition of *Hox* expression domains, which largely overlap in cephalochordates and vertebrates (ascidian data are still lacking). The motor of both events might be the decrease in size and transition to determinative development.

Methods

Whole-mount *in situ* hybridization

The ISH protocol was adapted from ref. 25 with the following modifications: embryos were fixed overnight in 4% paraformaldehyde in 0.1 M MOPS pH 7.5, 0.5 M NaCl at 4 °C, washed in 0.05 M Tris-HCl pH 8, treated for 1 min with 10 µg ml⁻¹ proteinase K in 0.05 M Tris-HCl pH 8 at 37 °C, followed by incubation in 1 M MOPS, 0.5 M NaCl, 0.1% Triton X-100 at 20 °C for 20 min. Embryos were prehybridized for 1 h and hybridized overnight in 50% deionized formamide, 5 × SSC, 1% blocking reagent at 60 °C. After the washing procedure, the hybridized embryos were blocked in 1% blocking reagent (Roche) and 1% filtered (0.45 µm pore size) lamb sera (Gibco) in PBS for 1 h at 20 °C. They were then incubated for a further 1 h at 20 °C in alkaline-phosphatase-coupled anti-digoxigenin Fab fragments (Roche) and in 1 × PBS, 0.1% Triton X-100 and treated for standard detection as described by Roche. Alternatively, the embryos were incubated in horseradish-peroxidase-conjugated anti-digoxigenin Pod fragments (Roche) for 4 h at 20 °C, followed by 1–4 days of incubation at 20 °C with the Tyramide Signal Amplification kit (Perkin Elmer) for fluorescence staining. Nuclear staining was obtained by incubation overnight at 4 °C with To-Pro-3 iodide (Molecular Probes). Specimens were mounted in Vectashield mounting medium (Vector Laboratories) and analysed with a Leica TCS laser scanning confocal microscope.

Received 13 March; accepted 3 June 2004; doi:10.1038/nature02709.

- Wada, H. Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny. *Mol. Biol. Evol.* **15**, 1189–1194 (1998).
- Finnerty, J. R. The origins of axial patterning in the metazoa: how old is bilateral symmetry? *Int. J. Dev. Biol.* **47**, 523–529 (2003).
- Balavoine, G., de Rosa, R. & Adoutte, A. *Hox* clusters and bilaterian phylogeny. *Mol. Phylogenet. Evol.* **24**, 366–373 (2002).
- Von Allmen, G. *et al.* Splits in fruitfly *Hox* gene complexes. *Nature* **380**, 116 (1996).
- Burglin, T. R. & Ruvkun, G. The *Caenorhabditis elegans* homeobox gene cluster. *Curr. Opin. Genet. Dev.* **3**, 615–620 (1993).
- Akam, M. *Hox* and *HOM*: homologous gene clusters in insects and vertebrates. *Cell* **57**, 347–349 (1989).
- Ferrier, D. E., Minguillon, C., Holland, P. W. & Garcia-Fernandez, J. The amphioxus *Hox* cluster: deuterostome posterior flexibility and *Hox14*. *Evol. Dev.* **2**, 284–293 (2000).
- Powers, T. P. & Amemiya, C. T. Evidence for a *Hox14* paralog group in vertebrates. *Curr. Biol.* **14**, R183–R184 (2004).
- Dehal, P. *et al.* The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157–2167 (2002).
- Spagnuolo, A. *et al.* Unusual number and genomic organization of *Hox* genes in the tunicate *Ciona intestinalis*. *Gene* **309**, 71–79 (2003).
- Seo, H. C. *et al.* Miniature genome in the marine chordate *Oikopleura dioica*. *Science* **294**, 2506 (2001).
- Boutanaev, A. M., Kalmykova, A. I., Shevelov, Y. Y. & Nurminsky, D. I. Large clusters of co-expressed genes in the *Drosophila* genome. *Nature* **420**, 666–669 (2002).
- Roy, P. J., Stuart, J. M., Lund, J. & Kim, S. K. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* **418**, 975–979 (2002).
- Gionti, M. *et al.* *Cihox5*, a new *Ciona intestinalis* *Hox*-related gene, is involved in regionalization of the spinal cord. *Dev. Genes Evol.* **207**, 515–523 (1998).
- Locascio, A. *et al.* Patterning the ascidian nervous system: structure, expression and transgenic analysis of the *CiHox3* gene. *Development* **126**, 4737–4748 (1999).
- Nagatomo, K. & Fujiwara, S. Expression of *Raldh2*, *Cyp26* and *Hox-1* in normal and retinoic acid-treated *Ciona intestinalis* embryos. *Gene Expr. Patterns* **3**, 273–277 (2003).
- Welsch, U. & Storch, V. Zur Feinstruktur der Chorda dorsalis niederer Chordaten Dendrodoa grossularia (v. Beneden) und Oikopleura dioica (Fol.). *Z. Zellforsch. Mikrosk. Anat.* **93**, 547–559 (1969).
- Lacalli, T. C. Tunicate tails, stolons, and the origin of the vertebrate trunk. *Biol. Rev. Camb. Phil. Soc.* **74**, 177–198 (1999).
- Ferrier, D. E. & Holland, P. W. *Ciona intestinalis* *ParaHox* genes: evolution of *Hox/ParaHox* cluster integrity, developmental mode, and temporal colinearity. *Mol. Phylogenet. Evol.* **24**, 412–417 (2002).
- van der Hoeven, F., Zakany, J. & Duboule, D. Gene transpositions in the *HoxD* complex reveal a hierarchy of regulatory controls. *Cell* **85**, 1025–1035 (1996).
- Averof, M. & Akam, M. *Hox* genes and the diversification of insect and crustacean body plans. *Nature* **376**, 420–423 (1995).
- Kmita, M. & Duboule, D. Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301**, 331–333 (2003).
- Cowing, D. & Kenyon, C. Correct *Hox* gene expression established independently of position in *Caenorhabditis elegans*. *Nature* **382**, 353–356 (1996).
- Aboobaker, A. A. & Blaxter, M. L. *Hox* gene loss during dynamic evolution of the nematode cluster. *Curr. Biol.* **13**, 37–40 (2003).
- Spada, F. *et al.* Molecular patterning of the oikopleuric epithelium of the larvacean tunicate *Oikopleura dioica*. *J. Biol. Chem.* **276**, 20624–20632 (2001).
- Swofford, D. L. *PAUP* 4.0: Phylogenetic Analysis Using Parsimony (*and other methods)*, Version 4.0b10 (Sinauer, Sunderland, Massachusetts, 2002).
- Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
- Basbash, S. & Postlethwait, J. Brachyury (T) expression in embryos of a larvacean urochordate, *Oikopleura dioica*, and the ancestral role of T. *Dev. Biol.* **220**, 322–332 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Adoutte for advice throughout the course of this work; R. Aasland, M. Scharlt, T. Stach and E. Thompson for their comments at several stages of manuscript preparation; and the personnel of the *Oikopleura* culture facility for technical support. Funding was provided by the Research Council of Norway and the University of Bergen in the frame of the Sars Centre – EMBL partnership contract.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.C. (Daniel.Chourrout@sars.uib.no). BAC clone sequences have been deposited in GenBank under accession numbers AY449458–AY449462 and AY613855–AY613856.

Biological abnormality of impaired reading is constrained by culture

Wai Ting Siok¹, Charles A. Perfetti², Zhen Jin³ & Li Hai Tan^{1,4}

¹Cognitive Neuroscience Laboratory, Department of Linguistics, University of Hong Kong, Hong Kong, China

²Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

³MRI Division, Beijing 306 Hospital, Beijing 100101, China

⁴Laboratory of Neuropsychology, National Institute of Mental Health, NIH, Bethesda, Maryland 20892, USA

Developmental dyslexia is characterized by a severe reading problem in people who have normal intelligence and schooling^{1–3}. Impaired reading of alphabetic scripts is associated with dysfunction of left temporoparietal brain regions^{2–5}. These regions perform phonemic analysis and conversion of written symbols to phonological units of speech (grapheme-to-phoneme conversion); two central cognitive processes that mediate reading acquisition^{6–7}. Furthermore, it has been assumed that, in contrast to cultural diversities, dyslexia in different languages has a universal biological origin^{1,8}. Here we show using functional magnetic resonance imaging with reading-impaired Chinese children and associated controls, that functional disruption of the left middle frontal gyrus is associated with impaired reading of the Chinese language (a logographic rather than alphabetic writing system). Reading impairment in Chinese is manifested by two deficits: one relating to the conversion of graphic form (orthography) to syllable, and the other concerning orthography-to-semantics mapping. Both of these processes are critically mediated by the left middle frontal gyrus, which functions as a centre for fluent Chinese reading^{9–11} that coordinates and integrates various information about written characters in verbal and spatial working memory. This finding provides an insight into the fundamental pathophysiology of dyslexia by suggesting that rather than having a universal origin, the biological abnormality of impaired reading is dependent on culture.

Unlike alphabetic writing systems that follow a design principle of mapping graphemes (visual form) onto phonemes (minimal phonological units of speech), the Chinese logographic system maps graphic forms (characters) onto morphemes (meanings). The phonology of written Chinese is defined at the monosyllabic level, with no parts of a character corresponding to phonemes. For instance, in the English word ‘beech’ the ‘b’ corresponds to /b/, and the latter is a segment of the word. However, the Chinese character 理 is pronounced /li3/ (meaning ‘reason’, where the numeral refers to Chinese tone), and its phonetic component 里, located on the right (also pronounced /li3/, meaning ‘inside’), does not correspond to a piece of the word’s phonological form. Hence, Chinese writing does not allow the segmental analysis that is fundamental to

alphabetic systems¹², so the letter–sound conversion rules that exist for all alphabetic languages are impossible in Chinese. Visually, Chinese characters possess a number of intricate strokes that are packed into a square shape, often having their meanings suggested by visual configurations¹³.

The characteristics of written Chinese as a morphosyllabic system present a sharp contrast with the alphabetic system that provides the critical dimensions of universal comparisons. Whereas behavioural research has established that impaired reading of the English language is caused by phonological deficits (that is, a failure to form quality links between orthography and phonology), the characterization of reading impairments in Chinese might be different. Research has shown that reading in Chinese entails recurrent and parallel activations of three interconnected linguistic constituents: orthography, meaning and phonology^{14,15}. On this view, reading difficulty in Chinese evolves not only from a poor quality mapping of orthography to phonology, but also from a substandard connection between orthography and semantics. Therefore, cortical activation irregularity associated with impaired Chinese reading is thought to differ from that associated with impaired English reading.

We performed two functional magnetic resonance imaging (fMRI) experiments to test this hypothesis. Experiment 1 sought to identify brain regions showing greater activation in normal Chinese readers during orthography-to-phonology mapping than in impaired Chinese readers. We used a homophone judgement design in which 16 children (average age 11) judged whether two synchronously exposed Chinese characters had an identical pronunciation (Fig. 1a; see Methods). In the control condition, children decided whether a pair of characters had the same physical size (font size decision). This task controlled for activation owing to the visual-orthographic and incidental semantic processing of the linguistic stimuli in the experimental task, thus allowing us to identify neural correlates of phonological processing and orthography-to-phonology mapping^{10,16}. Analysis of variance revealed that normal readers performed better than impaired readers in homophone judgements (response accuracy, $F_{1,13} = 4.744$, $P < 0.05$, and reaction time, $F_{1,13} = 8.258$, $P < 0.05$; Fig. 1d, e), but accuracy and

reaction time differences between homophone and font size decisions were not significant across normal and impaired readers (F -values < 1). This indicates that our neuroimaging results from comparisons of task and control conditions in the two groups were not due to performance differences⁶ or task difficulty¹⁷.

During homophone judgement (contrasted with font size decision), normal readers activated neural circuits involving left mid-inferior frontal gyrus, cingulate cortex and several regions in the occipital lobe (Table 1 and Fig. 2a), with 100% intersubject consistency particularly in mid-inferior frontal cortex and cingulate cortex (Fig. 2b). Impaired readers activated neural networks that essentially overlap those activated by normal readers, showing 100% intersubject consistency in left inferior prefrontal gyrus. Notably, direct comparisons of blood-oxygen-level-dependent (BOLD) activity of the two groups indicated that cortical activity was stronger in left middle frontal gyrus (LMFG) at Brodmann area (BA) 9 ($x = -50$, $y = 11$, $z = 34$) for normal readers than for impaired readers (Fig. 2c). Conversely, in left inferior prefrontal gyrus (BA 45: $x = -53$, $y = 23$, $z = 16$), impaired readers had stronger brain activation than normal readers (Fig. 2d). A correlation analysis of BOLD activation and children's reading performance indicated a significant positive correlation in LMFG ($r = 0.529$, $P < 0.05$) and a significant negative correlation in left inferior frontal gyrus ($r = -0.749$, $P < 0.001$), lending support to the results from group comparisons. Impaired readers showed activity in the left inferior parietal cortex, whereas normal readers

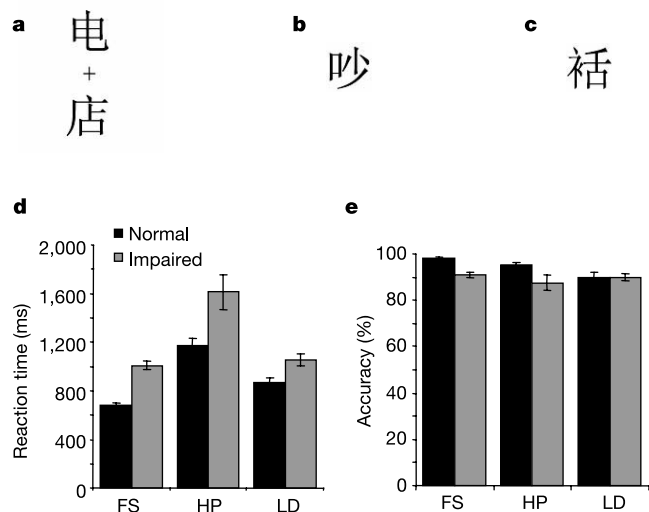


Figure 1 Examples of experimental materials and behavioural results. **a**, Two homophonic characters (pronounced /dian/). **b**, A real Chinese character. **c**, A non-character. From these examples, it can be seen that Chinese characters are composed of a number of strokes and are salient as holistic square-shaped units. **d**, Reaction time. **e**, Accuracy. Error bars indicate standard error measurement (s.e.m.). FS, font size judgement; HP, homophone decision; LD, lexical decision.

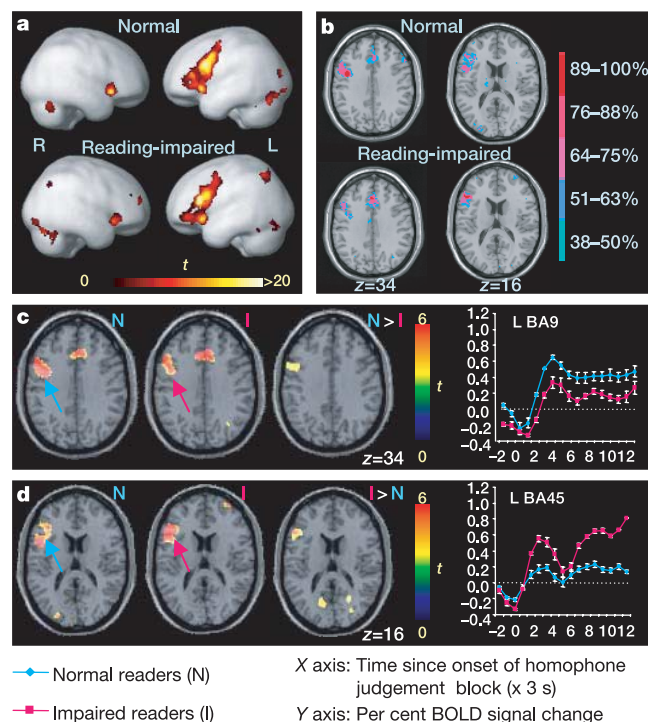


Figure 2 Brain regions with significant activity during phonological processing and orthography-to-phonology mapping. **a**, Cortical activation associated with homophone judgement contrasted with font size decision in normal and reading-impaired Chinese readers. **b**, Intersubject consistency. LMFG activity was found in all normal readers, whereas left inferior prefrontal activity was found in all impaired readers. **c**, **d**, Brain regions showing group differences during homophone judgement. Presented are time course and activation maps in LMFG (**c**) and left inferior frontal gyrus (**d**) as indicated by blue (normal readers) and red (impaired readers) arrows. Per cent BOLD signal change ($\pm$ s.e.m.) at voxels of maximal difference between groups is shown from the interval of -6 – 36 s. $N > I$, normal readers minus impaired readers; $I > N$, impaired readers minus normal readers.

Table 1 **Coordinates of activation peaks**

Brain area	Homophone judgement					Lexical decision				
	BA	Coordinates			t-score	BA	Coordinates			t-score
		X	Y	Z			X	Y	Z	
Normal readers										
Frontal lobe										
L middle frontal gyrus	9	-42	14	33	11.45	9	-51	10	38	14.11
	6	-46	0	48	8.38					
L inferior frontal gyrus	44	-48	11	23	11.96	44	-44	3	29	24.19
L insula						-	-32	25	1	13.84
R middle frontal gyrus						9	53	27	28	13.91
						6	40	-3	50	8.97
R inferior frontal gyrus	47	38	17	-8	8.14	44	46	9	25	20.73
						47	34	21	-4	15.79
Anterior cingulate gyrus	32	8	27	30	6.83	32	2	20	37	16.95
Medial frontal gyrus	8	-2	18	47	12.21	8	-2	30	41	19.10
Occipital lobe										
L fusiform gyrus	19	-34	-73	-13	6.39	19/37	-46	-65	-10	27.41
	37	-34	-55	-12	5.47	37	-38	-53	-18	18.65
L mid-inferior occipital	18	-26	-91	0	6.09	18	-28	-89	-1	19.34
	19	-26	-85	19	5.81					
L lingual gyrus	18	-28	-76	-8	6.29					
Parietal lobule										
R inferior parietal lobule						40	48	-29	46	7.33
Subcortical regions										
L thalamus						-	-10	-19	6	10.79
R thalamus						-	6	-19	6	9.71
						-	4	-10	4	9.18
L cerebellum	-	32	-63	-22	6.91	-	-32	-74	-38	6.21
Reading-impaired readers										
Frontal lobe										
L middle frontal gyrus	9	-38	6	38	7.35	9	-42	11	30	12.35
						10	-30	40	18	7.56
						46	-51	26	24	5.50
						9	-53	6	37	5.34
L inferior frontal gyrus	45	-51	26	13	12.07	47	-32	19	-1	13.29
	44/45	-42	13	21	11.75					
	47	-30	25	-5	11.73					
L precentral gyrus						4	-36	-21	54	8.73
						4	-50	-19	49	8.58
						4	-34	-28	62	5.15
						6	-32	-2	31	6.70
R superior frontal gyrus	10	32	57	14	5.98					
R middle frontal gyrus	10	22	44	20	5.70	10	38	53	7	11.11
						46	44	25	21	9.99
						6	34	2	50	8.24
R inferior frontal gyrus	47	32	25	-6	10.28	47	36	21	-8	15.20
	45	34	22	8	5.65					
Anterior cingulate gyrus	32	-2	19	40	13.04	32	4	23	34	17.42
	32	10	26	21	6.59					
Medial frontal gyrus	8	-2	18	47	12.60	6	-4	-9	48	6.86
Occipital lobe										
L superior occipital gyrus						19	-24	-70	31	9.53
L fusiform gyrus	37	-38	-51	-18	6.16	37/18	-40	-76	-11	17.12
L lingual gyrus	18	-6	-80	-14	6.40	18	-24	-93	0	18.65
L middle occipital gyrus	19	-40	-72	-8	5.65	19	-38	-72	-3	16.80
L cuneus	17	-16	-65	12	5.68					
	18	-12	-73	15	4.87					
R superior occipital gyrus	19	34	-62	40	5.84					
R fusiform gyrus						37	40	-55	-17	17.64
R mid-inferior occipital	18	32	-88	-6	6.30	18	38	-80	1	16.59
						19	36	-79	8	15.67
R lingual gyrus	19	30	-56	1	5.45					
Parietal lobule										
L inferior parietal lobule	40/19	-30	-62	40	6.51	40/7	-26	-58	42	11.77
L superior parietal lobule	7	-28	-64	47	6.50	7	-34	-52	56	8.66
R inferior parietal lobule						40	48	-31	42	6.52
Subcortical regions										
L thalamus	-	-12	-15	8	6.17	-	-28	-31	-3	8.28
R thalamus						-	6	-17	12	8.77
L cerebellum	-	-8	-85	-23	6.20					
	-	-30	-61	-17	5.55					
R cerebellum	-	8	-77	-21	8.45					
Normal > impaired readers										
Frontal lobe										
L middle frontal gyrus	9	-50	11	34	4.16	9	-50	10	38	18.02
L inferior frontal gyrus						44	-44	3	29	15.03
R middle frontal gyrus						9	50	29	30	4.23
R inferior frontal gyrus						44	44	5	26	11.69
Occipital lobe										
L fusiform gyrus						19/37	-46	-65	-12	11.64
Impaired > normal readers										
Frontal lobe										
L inferior frontal gyrus	45	-53	23	16	12.07					
Occipital lobe										
R inferior occipital cortex						18	26	-78	-3	8.15

L, left; R, right; normal > impaired readers, normal readers minus impaired readers; impaired > normal readers, impaired readers minus normal readers.

did not (Fig. 2a). However, the between-group comparison did not reveal this difference to be statistically significant, owing to low intersubject consistency (25%) of the activation of this brain area in impaired readers.

These results suggest that there is an important neurological difference with respect to impaired Chinese and English reading. Previous research using similar designs has consistently revealed reduced activation in left temporoparietal regions as a biological signature of English reading disability^{2,4,5}. Furthermore, it has been shown that behavioural remediation ameliorates these dysfunctional neural mechanisms in American children with dyslexia^{3,5}. In experiment 1, we observed reduced activation of LMFG, instead of left temporoparietal regions, as a neuro-anatomical marker of Chinese reading difficulty. Our findings agree with previous imaging data showing that LMFG is a crucial brain area responsible for skilled Chinese reading^{9–11}. Broadly speaking, the LMFG carries out the representation and working memory of visuo-spatial and verbal information, and coordinates cognitive resources as a central executive system^{18–20}. Chinese characters are composed of intricate strokes packed into a square shape; recognizing the character demands detailed visuo-spatial computation and, during homophone judgement, a high inter-activity of orthography and phonology. The LMFG is recruited to serve these functions.

We also noted the hyperactivity of the anterior portion of the left

inferior prefrontal cortex (BA 45) arising during impaired Chinese reading. This finding is compatible with findings of English reading impairment⁴, suggesting that impaired Chinese reading engages neural circuits in anterior regions to compensate for the disruption in mid-dorsal frontal systems.

To determine whether brain regions mediating orthography-to-semantics mapping show abnormal activity in impaired Chinese reading, we performed a second experiment using the same subjects in a character decision design that required children to judge whether a viewed stimulus was a real Chinese character (Fig. 1b, c). This task focuses on the processing of visual-orthographic and semantic constituents and their interconnections (see Methods), although phonological information of 'legal' characters may also be relevant. Behavioural results indicated that normal readers performed better than impaired readers in reaction time (Fig. 1d; $F_{1,14} = 9.14$, $P < 0.01$), but not in response accuracy (Fig. 1e; $F_{1,14} = 0.004$, $P = 0.949$). Brain activations related to character decision (contrasted with fixation baseline) were seen bilaterally in mid-inferior frontal lobes, inferior parietal cortex and occipital cortex (Fig. 3a and Table 1), with very high intersubject consistency particularly in middle and inferior frontal gyri and visual cortex (Fig. 3b). Nevertheless, important differences in the magnitude of cortical activity were observed between normal and reading-impaired children (Fig. 3c–h). Impaired readers showed weaker activation than normal readers in the following areas: bilateral

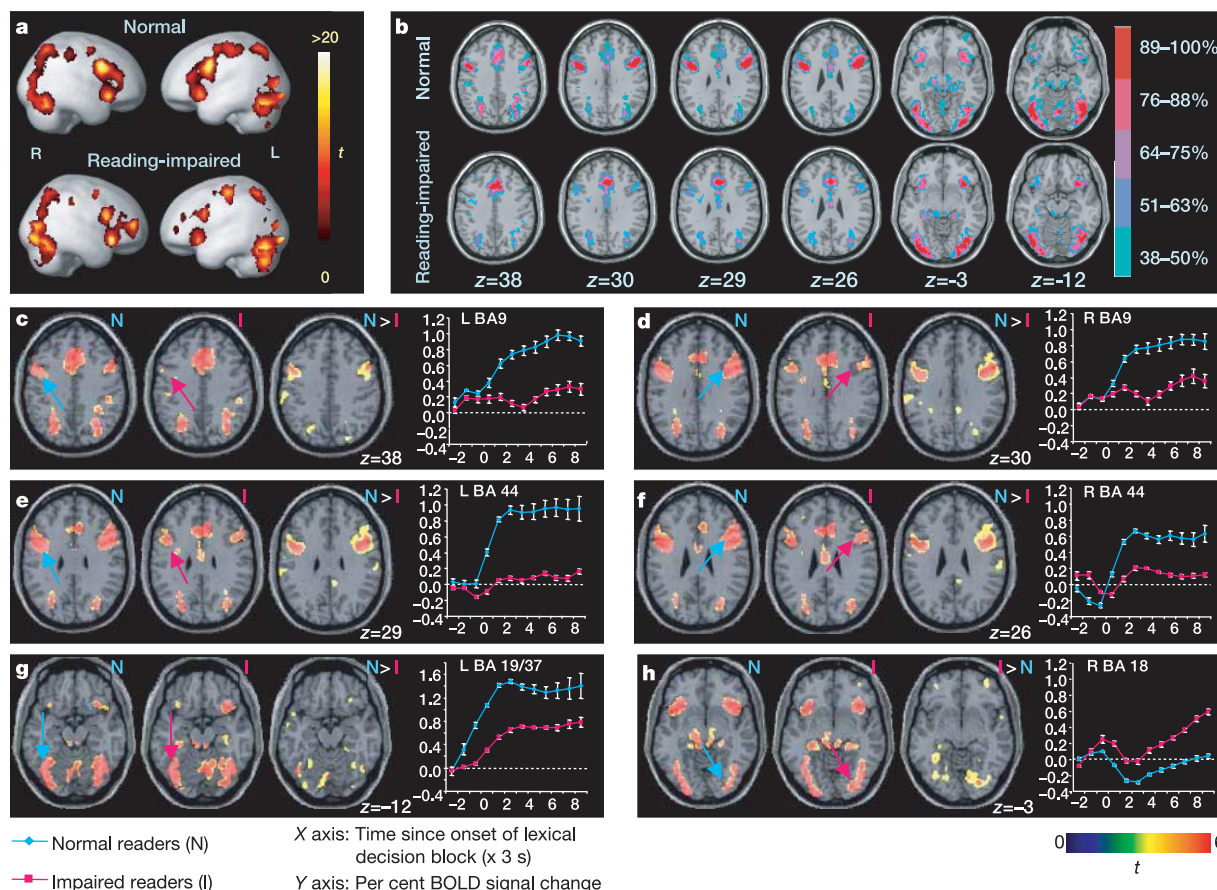


Figure 3 Brain regions with significant activity during orthography-to-semantics mapping. **a**, Cortical activation associated with Chinese character decision (contrasted with fixation). **b**, Intersubject consistency. **c–h**, Group differences during character decision. Presented are time course and activation maps in LMFG, **c**; right MFG, **d**; left inferior frontal gyrus, **e**; right inferior frontal cortex, **f**; left fusiform gyrus, **g**; and right

inferior occipital cortex, **h**, as indicated by blue (normal readers) and red (impaired readers) arrows, respectively. Per cent BOLD signal change ($\pm$ s.e.m.) at voxels of maximal difference between groups is shown from the interval of -6 – 30 s. $N > I$, normal readers minus impaired readers; $I > N$, impaired readers minus normal readers.

middle frontal gyrus at BA 9 ($x = -50$, $y = 10$, $z = 38$ on the left, and $x = 50$, $y = 29$, $z = 30$ on the right), bilateral inferior prefrontal gyrus at BA 44 ($x = -44$, $y = 3$, $z = 29$ on the left, and $x = 44$, $y = 5$, $z = 26$ on the right), and left fusiform gyrus at BA 19/37 ($x = -46$, $y = -65$, $z = -12$). Impaired readers showed greater activation than controls in right inferior occipital cortex near BA 18 ($x = 26$, $y = -78$, $z = -3$). Correlation analyses of BOLD activation and reading performance revealed significant correlations in these brain regions (in left BA 9, $r = 0.745$, $P < 0.001$; right BA 9, $r = 0.707$, $P < 0.002$; left BA 44, $r = 0.830$, $P < 0.001$; right BA 44, $r = 0.814$, $P < 0.001$; left BA 19/37, $r = 0.822$, $P < 0.001$; and right BA 18, $r = -0.775$, $P < 0.001$).

Experiment 2 thus demonstrated a failure of large neural circuits to function properly during impaired Chinese reading. The reduced activation of LMFG agrees with the finding of experiment 1, suggesting that this brain region coordinates and integrates visual-orthographic and semantic (and phonological) processes in verbal and spatial working memory. This proposal is corroborated by our finding of greater activation in normal readers in left fusiform gyrus, a visual word form area that serves orthographic-to-semantic integration both in English²¹ and Chinese²². The stronger activity of the posterior portion of the left inferior frontal gyrus for normal readers implies that this cortical region is particularly relevant to orthographic-semantic processing and also that controlled meaning retrieval^{23,24} may mediate lexical decisions. The strong activity of left fusiform and inferior frontal systems in normal Chinese reading seen in experiment 2 is in line with imaging studies of English reading development⁶ and reading remediation³.

Two regions in the right hemisphere showed differences between normal and impaired Chinese readers. Normal readers showed stronger activation in right mid-inferior frontal gyrus, whereas impaired readers showed stronger activation in right inferior occipital cortex. Right mid-inferior frontal regions contribute to fluent Chinese reading⁹, and right inferior occipital cortex is thought to be engaged in visual processing of Chinese characters^{22,25}. The increased activity of right inferior occipital gyrus in our experiment indicates that Chinese impaired readers struggle even in the visuo-spatial analysis of printed characters.

These two experiments provide compelling evidence that Chinese reading disability is characterized by the dysfunction of neural circuits responsible for the mapping of orthography-to-phonology and orthography-to-semantics. This pattern of findings is important for several reasons. First, it reveals a significant variation between the neural substrates of Chinese and English impaired reading. The LMFG is crucial to normal Chinese reading, as its dysfunction is associated with Chinese reading difficulty. Our findings support the idea that cognitive strategies for reading development tune the cortex²⁶. This idea draws a parallel with a recent proposal that anatomical brain differences can be produced by differences in cultures, as suggested by a localized morphological analysis of brain anatomy that found that LMFG is anatomically larger in Chinese-speaking Asians than in English-speaking Caucasians²⁷. Second, our fMRI findings showed that Chinese reading entails coherent and orchestrated activities of key neuroanatomical modules relating to orthography, semantics and phonology. This lends support to the assumption of parallel activation of meaning and phonology made by the interactive constituency model of (Chinese) reading^{14,15}, while also posing a major challenge to the biological unit theory of dyslexia⁸. Finally, these data suggest that cross-cultural studies require great care to offer key dimensions of universal comparisons. Contrasts between writing systems with fundamental differences in their design principles, such as English and Chinese, allow understanding of both universal

and particular facets of the biological basis of reading disability. □

Methods

Sixteen children underwent fMRI scans: eight impaired readers (six boys and two girls, mean age = 10 yr 11 months, range 10 yr 2 months to 12 yr 5 months) and eight normal reader controls (four boys and four girls, mean age = 11 yr 1 month, range 10 yr 5 months to 12 yr 4 months). The children were fourth and fifth graders from the Yuquan Primary School in Beijing, and were physically healthy and free of neurological disease, head injury and psychiatric disorder. All subjects participated in experiments 1 and 2. Because there is no standardized reading ability test in China, the classification of children's reading performance was based primarily on their school performance in the Chinese reading course and their teacher's recommendation. For impaired readers, reading scores in the school fell below the 5th percentile among all children in the same grade. Children defined as normal readers had reading scores above the 60th percentile. In addition, to assess reading ability and non-verbal IQ, a commonly used (but not standardized) reading test²⁸ comprising 120 Chinese characters and the Raven IQ test were administered to all 400 children in the fourth and fifth grades. The reading performance of these children was distributed normally, with an average score of 71 (s.d. = 14). The eight normal and eight reading-impaired subjects for our experiments had normal and matched non-verbal intelligence, with an average Raven IQ falling in the 92nd percentile for both groups. The reading-impaired group had a mean character reading score of 42, two standard deviations below the average and significantly discrepant from their normal, non-verbal intelligence. The two groups of children were significantly different in reading performance ($t = 32.901$, $P < 0.001$). All children were native speakers of Putonghua, the official dialect of mainland China and the language of instruction in school. They were strongly right-handed as judged by the handedness inventory devised by Snyder and Harris²⁹. Informed consent was obtained from each subject and their parents before testing.

Design and materials

Experiment 1 used a phonological processing task; that is, homophone judgement. Experiment 2 used a character decision design (Fig. 1). The non-characters used in 'No' trials in experiment 2 were orthographically legal but without meanings or pronunciations (Fig. 1c). To perform the character decision successfully, children must rely on visual-orthographic familiarity and semantic accessibility of a viewed stimulus. All Chinese characters used in the two experiments were commonly encountered and selected from Chinese language textbooks of primary school grades 1 to 3. The visual complexity of the characters was matched across all conditions.

Each experiment was performed in a separate scan within the same session. In the homophone judgement scan, three blocks of homophone judgements were alternated with three blocks of font size judgements, which served as the baseline. Each block consisted of a 6-s instruction and 15 trials. In each trial, a pair of characters was synchronously exposed for 2,000 ms, one above and one below a fixation crosshair, followed by a 1,000-ms blank interval. In the character decision scan, four blocks of lexical decision task were alternated with four blocks of baseline task (visual fixation). Each task block consisted of a 6-s instruction and 20 trials. In each trial, a stimulus was displayed for 1,300 ms, followed by a 200-ms blank interval. For both experiments, subjects indicated a positive response by pressing the key corresponding to the index finger of their right (dominant) hand and a negative response by pressing the key corresponding to the index finger of their left (non-dominant) hand. They were asked to perform the tasks as quickly and accurately as possible. The behavioural data from one normal reader were not recorded correctly during the homophone judgement task and so were excluded from our statistical analysis of behavioural performance.

MRI acquisition

MRI scans were performed on a 2 T GE/Escint Prestige MRI scanner at Beijing 306 Hospital. Visual stimuli were presented to subjects through a projector onto a translucent screen. Subjects viewed the stimuli through a mirror attached to the head coil. A T_2^* -weighted gradient-echo echo planar imaging (EPI) sequence was used for fMRI scans, with the slice thickness = 6 mm, in-plane resolution = 2.9×2.9 mm and repetition time/echo time/flip angle = 3,000 ms/45 ms/90°. Twenty contiguous axial slices were acquired parallel to the anterior commissure-posterior commissure (AC-PC) line covering the whole brain. High-resolution ($2 \times 2 \times 2$ mm³) anatomical images were acquired using a T_1 -weighted, three-dimensional gradient-echo sequence.

Data analysis

SPM99 was used for image pre-processing and statistical analyses (<http://www.fil.ion.ucl.ac.uk/spm>). The functional images were realigned and normalized to an EPI template based on the ICBM152 stereotactic space (an approximation of canonical space). They were re-sampled into $2 \times 2 \times 2$ -mm cubic voxels and spatially smoothed by an isotropic gaussian kernel (6 mm full width at half-maximum). Activation maps were generated by using the general linear model in which time series were modelled using a boxcar function for every condition and convolved with the canonical haemodynamic response function. Adjusted mean images were created for each condition after removing global signal and low-frequency covariates. For each group, contrast images between task and baseline conditions were computed using the t statistic, which generated statistical parametric maps of t -values. The voxelwise threshold was set at $P < 0.05$ corrected for multiple comparisons, with an extent threshold of 20 contiguous voxels. Differences in activation maps between the two reading groups were examined with a two-sample t -test, and statistical threshold was set at $P < 0.001$ uncorrected, with an extent threshold of

20 contiguous voxels. Brain regions were estimated from Talairach and Tournoux³⁰, after adjustments for differences between MNI and Talairach coordinates. To evaluate the intersubject consistency of brain activations associated with normal and impaired reading in Chinese, we created penetrance maps by combining binary individual functional maps³¹. Penetrance maps were then overlaid on axial views of T_1 -weighted images to demonstrate the voxels with significant activation in three or more subjects. The binary functional maps were determined using $P < 0.05$ corrected for each subject.

Received 5 April; accepted 13 July 2004; doi:10.1038/nature02865.

- Eden, G. & Moats, L. The role of neuroscience in the remediation of students with dyslexia. *Nature Neurosci.* **5**, 1080–1084 (2002).
- Horwitz, B., Rumsey, J. M. & Donohue, B. C. Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc. Natl Acad. Sci. USA* **95**, 8939–8944 (1998).
- Temple, E. *et al.* Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from functional MRI. *Proc. Natl Acad. Sci. USA* **100**, 2860–2865 (2003).
- Shaywitz, S. E. *et al.* Functional disruption in the organization of the brain for reading in dyslexia. *Proc. Natl Acad. Sci. USA* **95**, 2636–2641 (1998).
- Aylward, E. H. *et al.* Instructional treatment associated with changes in brain activation in children with dyslexia. *Neurology* **61**, 212–219 (2003).
- Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. & Eden, G. Development of neural mechanisms for reading. *Nature Neurosci.* **6**, 767–773 (2003).
- Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D. & Seidenberg, M. S. How psychological science informs the teaching of reading. *Psychol. Sci. Publ. Interest* **2**, 31–74 (2001).
- Paulesu, E. *et al.* Dyslexia: cultural diversity and biological unity. *Science* **291**, 2165–2167 (2001).
- Tan, L. H. *et al.* The neural system underlying Chinese logograph reading. *Neuroimage* **13**, 836–846 (2001).
- Tan, L. H. *et al.* Neural systems of second language reading are shaped by native language. *Hum. Brain Mapp.* **18**, 158–166 (2003).
- Siok, W. T., Jin, Z., Fletcher, P. & Tan, L. H. Distinct brain regions associated with syllable and phoneme. *Hum. Brain Mapp.* **18**, 201–207 (2003).
- Mattingly, I. G. in *Language by Ear and by Eye: The Relationships Between Speech and Reading* (eds Kavanagh, J. F. & Mattingly, I. G.) 133–147 (MIT Press, Cambridge, Massachusetts, 1972).
- Wang, W. S.-Y. The Chinese language. *Sci. Am.* **228**, 50–62 (1973).
- Perfetti, C. A. & Tan, L. H. in *Reading Chinese Script: A Cognitive Analysis* (eds Wang, J., Inhoff, A. W. & Chen, L.) 115–134 (Lawrence Erlbaum Associates, Mahwah, New Jersey, 1999).
- Perfetti, C. A., Liu, Y. & Tan, L. H. The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychol. Rev.* (in the press).
- Price, C. J., More, C. J., Humphreys, G. W. & Wise, R. S. J. Segregating semantic from phonological processes during reading. *J. Cogn. Neurosci.* **9**, 727–733 (1997).
- Gabrieli, J. D., Poldrack, R. A. & Desmond, J. E. The role of left prefrontal cortex in language and memory. *Proc. Natl Acad. Sci. USA* **95**, 906–913 (1998).
- Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G. & Haxby, J. V. An area specialized for spatial working memory in human frontal cortex. *Science* **279**, 1347–1351 (1998).
- D'Esposito, M. *et al.* The neural basis of the central executive system of working memory. *Nature* **378**, 279–281 (1995).
- Petrides, M., Alivisatos, B., Meyer, E. & Evans, A. C. Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proc. Natl Acad. Sci. USA* **90**, 878–882 (1993).
- Cohen, L. *et al.* The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123**, 291–307 (2000).
- Tan, L. H. *et al.* Brain activation in the processing of Chinese characters and words: A functional MRI study. *Hum. Brain Mapp.* **10**, 16–27 (2000).
- Gold, B. T. & Buckner, R. L. Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron* **35**, 803–812 (2002).
- Wagner, A. D., Pare-Blagoev, J. E., Clark, J. & Poldrack, R. A. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* **31**, 329–338 (2001).
- Fu, S., Chen, Y., Smith, S., Iversen, S. & Matthews, P. M. Effects of word form on brain processing of written Chinese. *Neuroimage* **17**, 1538–1548 (2002).
- Fiez, J. A. Sound and meaning: how native language affects reading strategies. *Nature Neurosci.* **3**, 3–5 (2000).
- Kochunov, P. *et al.* Localized morphological brain differences between English speaking Caucasian and Chinese speaking Asian populations: New evidence of anatomical plasticity. *Neuroreport* **14**, 961–964 (2003).
- Siok, W. T. & Fletcher, P. The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Dev. Psychol.* **36**, 887–899 (2001).
- Snyder, P. J. & Harris, L. J. Handedness, sex, and familial sinistrality effects on spatial tasks. *Cortex* **29**, 115–134 (1993).
- Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme Medical Publishers, New York, 1988).
- Fox, P. T. *et al.* A PET study of the neural system of stuttering. *Nature* **382**, 158–161 (1996).

Acknowledgements This work was supported by a Hong Kong Government RGC Central Allocation grant, and by the Intramural Research Program of NIMH. We thank J. Gabrieli, R. Desimone, C. K. Leong, K. K. Luke, S. Matthews and J. Spinks for support and comments.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.H.T. (lihaitan@intr.nimh.nih.gov).

Reading the Hedgehog morphogen gradient by measuring the ratio of bound to unbound Patched protein

Andreu Casali & Gary Struhl

Howard Hughes Medical Institute, Department of Genetics and Development, Columbia University, New York, New York 10032, USA

Morphogens are 'form-generating' substances that spread from localized sites of production and specify distinct cellular outcomes at different concentrations. A cell's perception of morphogen concentration is thought to be determined by the number of active receptors, with inactive receptors making little if any contribution¹. Patched (Ptc)^{2–5}, the receptor for the morphogen Hedgehog (Hh)^{6–12}, is active in the absence of ligand and blocks the expression of target genes by inhibiting Smoothened (Smo), an essential transducer of the Hh signal^{3,13–16}. Hh binding to Ptc abrogates the ability of Ptc to inhibit Smo, thereby unleashing Smo activity and inducing target gene expression^{2,3,12–16}. Here, we show that a cell's measure of ambient Hh concentration is not determined solely by the number of active (unliganded) Ptc molecules. Instead, we find that Hh-bound Ptc can titrate the inhibitory action of unbound Ptc. Furthermore, we demonstrate that this effect is sufficient to allow normal reading of the Hh gradient in the presence of a form of Ptc that cannot bind the ligand¹² but retains its ability to inhibit Smo. These results support a model in which the ratio of bound to unbound Ptc molecules determines the cellular response to Hh.

Hh signal transduction is unusual in that the unbound receptor (Ptc) is the active form (that is, it keeps the pathway switched 'off' by inhibiting the transducer (Smo)). Binding by the ligand (Hh) inactivates the receptor, releasing Smo from inhibition and turning the pathway on^{2,3,12–16}. The cell's perception of the amount of ambient Hh has been proposed to be determined solely by the number of unliganded (active) Ptc molecules¹⁷. Thus, as Hh rises from nil to peak concentrations, Smo activity would increase merely as a consequence of the progressive depletion of the pool of active Ptc protein. According to this depletion model, liganded Ptc would be functionally equivalent to the absence of Ptc. Alternatively, liganded Ptc might titrate the inhibitory activity of unliganded Ptc so that a cell's perception of Hh concentration would depend on the ratio of the two forms¹².

To distinguish these possibilities, we have expressed different levels of constitutively active Ptc^{Δloop2} protein, a deleted form of Ptc (Fig. 1a) that cannot bind Hh but can still repress Smo¹², and asked whether the minimum amount of Ptc^{Δloop2} necessary to shut down the pathway depends on the presence of liganded Ptc. In the simple depletion model, in which Hh binding merely inactivates Ptc, Ptc^{Δloop2} should be impervious to liganded Ptc, and the minimum amount of Ptc^{Δloop2} required to turn off the pathway should not change. In the titration model, the presence of liganded Ptc should counteract the inhibitory activity of Ptc^{Δloop2}, requiring higher levels of Ptc^{Δloop2} to keep the pathway off.

Three Flp-out transgenes⁶ called $L > P^{Δ2}$, $M > P^{Δ2}$ and $H > P^{Δ2}$ (Fig. 1b; see Methods) were used to express low, medium or high levels of Ptc^{Δloop2} protein in clones of imaginal wing disc cells (Fig. 1b; see Methods). Wing discs are subdivided into anterior and posterior compartments, with cells in the posterior compartment programmed to express and those in the anterior compartment to respond to Hh^{6–9,18}. Ptc is normally expressed in anterior compartment cells only at low levels, except in those anterior cells that are located close to the anteroposterior compartment boundary

where Ptc expression is markedly upregulated by Hh^{2,7,8,16,19,20} (Fig. 1c). To quantify Ptc^{Δloop2} expression generated by each transgene, we measured the amount of Ptc staining in posterior compartment clones in the wing imaginal disc relative to the basal level of endogenous Ptc expressed in anterior cells away from the compartment boundary (defined as 100%; see Methods). $L > P^{Δ2}$, $M > P^{Δ2}$ and $H > P^{Δ2}$ clones in the posterior compartment expressed levels of Ptc^{Δloop2} that were approximately 80%, 140% and 320%, respectively, of the basal level (Fig. 1c), well within the physiological range of Ptc expression, which normally peaks at around 700% in anterior cells that abut the boundary.

To examine the consequences of expressing each transgene on Hh transduction, we assayed two target genes, *decapentaplegic* (*dpp*) and *collier* (*col*; also known as *knot*), which are induced by low and high thresholds of Hh signalling, respectively. *dpp*, monitored using a *dpp-lacZ* gene, is normally expressed in a broad stripe of

approximately 15–20 cell diameters that has a sharp peak at the compartment boundary and declines anteriorly as a function of distance from the posterior compartment^{6–8,12}. In contrast, *col*, monitored by the presence of Col protein, is expressed in a narrower band of around 5–10 cells²¹. $L > P^{Δ2}$ clones had no effect on Hh transduction (Fig. 2a), except at the extreme, anterior limit of the *dpp-lacZ* domain where *dpp-lacZ* expression was blocked in cells that would otherwise express barely detectable levels (data not shown). $M > P^{Δ2}$ clones showed a more readily detectable effect, generally reducing the level of *dpp-lacZ* and Col expression (Fig. 2b). By contrast, $H > P^{Δ2}$ clones were completely refractory to Hh: they failed to express either *dpp-lacZ* or Col even when located next to the compartment boundary (Fig. 2c).

To determine whether the same, high level of Ptc^{Δloop2} expression is necessary to block Hh pathway activity in the absence of endogenous Ptc, we generated clones of $L > P^{Δ2}$, $M > P^{Δ2}$ and $H > P^{Δ2}$ cells that were also *ptc*[−] (see Methods). *dpp-lacZ* and Col are constitutively expressed in *ptc*[−] cells; however, both responses were completely blocked when these cells expressed any one of the three transgenes (Fig. 2d and data not shown). This effect is particularly notable in the case of the $L > P^{Δ2}$ transgene because its expression has virtually no effect on Hh transduction in otherwise wild-type cells. Indeed, entirely $L > P^{Δ2}$ animals survive as phenotypically normal adults (Fig. 2e); however, $L > P^{Δ2}$ animals differ from wild-type animals in that Hh transduction is blocked, rather than constitutively activated, in clones of *ptc*[−] cells (Fig. 2e–g). This result confirms that Hh transduction in $L > P^{Δ2}$ cells depends on the presence of both Hh and endogenous Ptc: in the absence of either (for example, in anterior cells located away from the compartment boundary, or in *ptc*[−] cells located anywhere within the anterior compartment) the Hh transduction pathway is off. We therefore infer that in $L > P^{Δ2}$ cells liganded Ptc is not equivalent to the absence of Ptc; instead, it is responsible in this context for normal activation of the Hh pathway. This result indicates that liganded Ptc titrates the inhibitory activity of unliganded Ptc^{Δloop2}, and hence that it is the ratio of liganded to unliganded Ptc molecules rather than the absolute number of unliganded Ptc molecules that determines the Hh response.

As a further test of these conclusions, we examined $M > P^{Δ2}$ clones in the posterior compartment, where cells normally make and are exposed to uniformly high levels of Hh, and where Smo is constitutively active owing to the absence of endogenous Ptc^{22–24}. To assay Ptc activity in this context, we expressed a green fluorescent protein (GFP)-tagged, functionally inactive form of the transcription factor Cubitus interruptus (CiΔZnGFP; Fig. 3a). Ci is normally stabilized and activated in response to Hh^{25–27}. When CiΔZnGFP is expressed throughout the presumptive wing blade it is stabilized in all posterior compartment cells in response to the unfettered activity of Smo (Fig. 3a). Posterior compartment clones of $M > P^{Δ2}$ cells destabilize CiΔZnGFP (Fig. 3b), as expected because Ptc^{Δloop2} cannot bind Hh and constitutively inhibits Smo. However, when these clones carried a second transgene, $M > P^{+}$, that expresses Ptc⁺ in the posterior compartment³, the CiΔZnGFP protein is once again stabilized (Fig. 3c). Because Hh is normally expressed at peak levels in cells of the posterior compartment, we infer that some if not most of the ectopic Ptc⁺ protein expressed in the posterior compartment is bound by Hh. Hence, as in the anterior compartment, it seems that liganded Ptc can drive Hh transduction in the presence of constitutively active Ptc^{Δloop2} that would otherwise shut down the pathway.

To estimate the range of ratios between unliganded and liganded Ptc that may normally distinguish between off and on states of the Hh pathway, we performed a titration experiment in which we generated clones of *ptc*[−] cells expressing different ratios of constitutively unliganded Ptc (Ptc^{Δloop2}) and Hh–Ptc, a chimaeric protein that seems to behave like a constitutively liganded form of Ptc (Fig. 1a; see Methods).

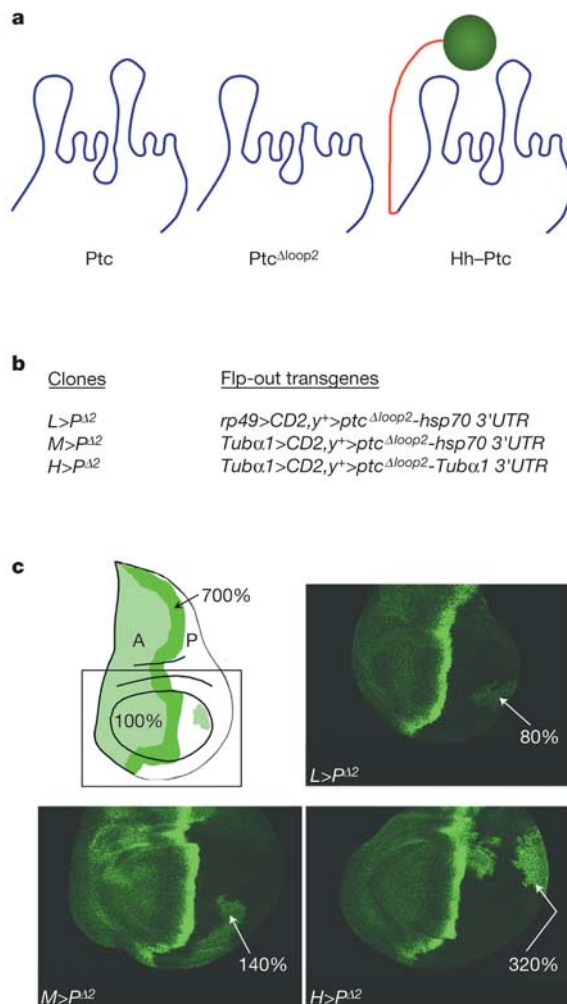


Figure 1 Generation of clones expressing low ($L > P^{Δ2}$), medium ($M > P^{Δ2}$) and high ($H > P^{Δ2}$) levels of Ptc^{Δloop2}. **a**, Ptc^{16,19,20} protein and the two modified versions, Ptc^{Δloop2} (ref. 12) and Hh–Ptc (see Methods), used in this study. **b**, Flip-out transgenes used to generate clones of $L > P^{Δ2}$, $M > P^{Δ2}$ and $H > P^{Δ2}$ cells by Flp-mediated recombination of the *>CD2,y⁺>* cassette (Methods). **c**, Wing imaginal discs bearing posterior compartment Flp-out clones of $L > P^{Δ2}$, $M > P^{Δ2}$ and $H > P^{Δ2}$ cells stained for Ptc protein. The average level of expression of $L > P^{Δ2}$, $M > P^{Δ2}$ and $H > P^{Δ2}$ was approximately 80% ($n = 6$; s.d. = 19%), 140% ($n = 9$; s.d. = 22%) and 320% ($n = 12$; s.d. = 41%), respectively, relative to the basal level of endogenous Ptc staining in anterior compartment cells away from the anteroposterior compartment boundary (see Methods). The peak level of endogenous Ptc staining in anterior cells abutting the compartment boundary was ~700% ($n = 12$; s.d. = 165%).

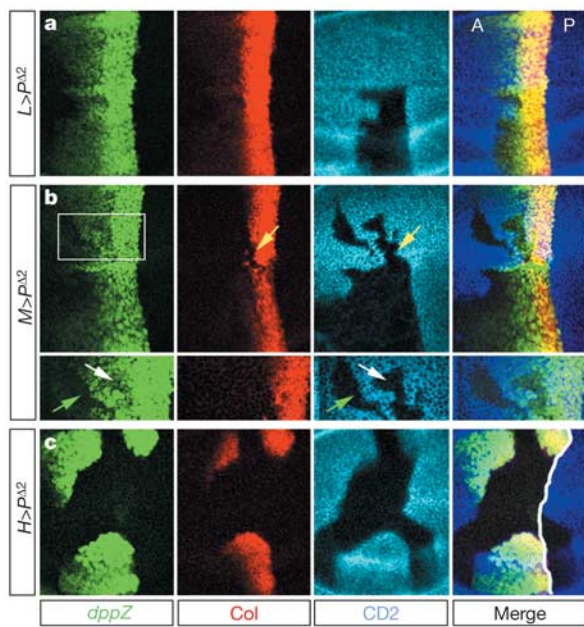
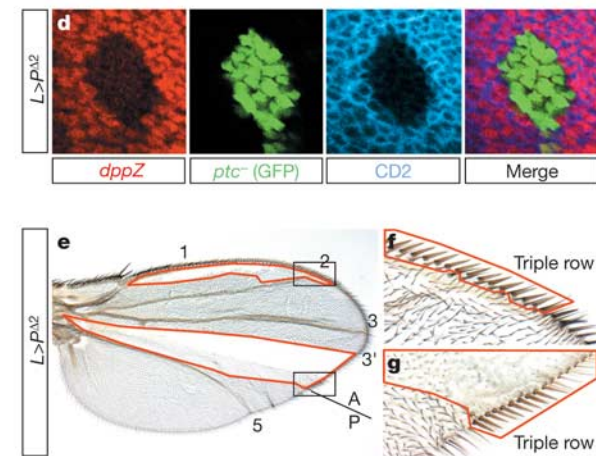


Figure 2 Different amounts of Ptc^{Δloop2} are required to inhibit the Hh pathway depending on the presence or absence of endogenous Ptc. **a–c**, Wing discs bearing Flip-out clones of $L > P^{\Delta 2}$ (**a**), $M > P^{\Delta 2}$ (**b**) and $H > P^{\Delta 2}$ (**c**) cells in the anterior compartment, marked black by the absence of CD2 staining (blue). The anteroposterior (A–P) compartment boundary is marked by the sharp border of *dpp-lacZ* (*dppZ*; green) and Collier (Col; red) expression in **a** and **b**, and by a white line in **c** (anterior to the left). Expression of *dpp-lacZ* and Col appear normal in $L > P^{\Delta 2}$ clones, but are reduced or abolished in $M > P^{\Delta 2}$ and $H > P^{\Delta 2}$ clones. In the boxed region of **b** (magnified in the bottom panels), white and green arrows mark, respectively, reduced or abolished expression of *dpp-lacZ*; the yellow arrow in the main panel marks the absence of Col expression. **d**, Low level Ptc^{Δloop2} expression ($L > P^{\Delta 2}$) in an anterior compartment clone of *ptc*[−] cells close to the compartment boundary (*ptc*[−] genotype marked by the presence of GFP (green)). The $L > P^{\Delta 2}$ genotype is marked black by the absence of CD2 (blue); *dpp-lacZ* expression is repressed. **e**, Two anterior compartment clones of *ptc*[−] cells (outlined in red) in a wing in which all cells carry the flipped-out $L > P^{\Delta 2}$ transgene (veins 1, 2, 3 and 5, as well as the



anteroposterior compartment boundary are indicated where they abut the wing margin); both clones behave as if mutant for *smo*. The anteriorly situated clone (top) is phenotypically wild type, indicating that the Hh transduction pathway is off. The clone along the anteroposterior compartment boundary (bottom) forms a mirror symmetric duplication of a more anterior pattern, including an ectopic vein 3 (3') and triple row bristles (**g**), at the expense of posterior compartment pattern (vein 4 is absent). As in the case of *smo*[−] clones, this distinctive phenotype reflects the failure of the mutant cells to transduce and sequester Hh, displacing the source of Dpp signalling to the middle of the anterior compartment (see refs 3, 12). **f, g**, Details of the boxed portions in **e**, showing, respectively, normal and ectopic 'triple row' bristles formed by *ptc*[−] cells, confirming that the Hh transduction pathway is off (clones are marked by yellow, which lightens bristle colour (**f, g**), and *shavenoid*, which removes hairs (**g**); the red outlines in **e** show the contribution of the clones to the dorsal surface; only those portions of the wing in which both the dorsal and ventral surfaces are mutant appear hairless in this image).

We generated two transgenes, $L > hh-ptc$ and $H > hh-ptc$, and quantified their expression level as 50% and 220% relative to endogenous Ptc in anterior cells not exposed to Hh (Fig. 4 legend). By activating different combinations of $L > P^{\Delta 2}$, $M > P^{\Delta 2}$, $L > hh-ptc$ and $H > hh-ptc$ transgenes in anterior compartment clones of *ptc*[−] cells, we generated cells in which the only Ptc proteins present are Hh–Ptc and Ptc^{Δloop2}, at ratios of 220%:80% ($H > hh-ptc:L > P^{\Delta 2}$), 220%:140% ($H > hh-ptc:M > P^{\Delta 2}$) or 50%:80% ($L > hh-ptc:L > P^{\Delta 2}$). Both *dpp-lacZ* and Col expression were fully activated at the highest ratio of Hh–Ptc:Ptc^{Δloop2} (220%:80%; Fig. 4a, b); however, neither was expressed in cells in which the ratio was reduced, either by an approximately twofold increase in the level of Ptc^{Δloop2} (220%:140%; Fig. 4c and data not shown), or by an approximately fourfold decrease in the level of Hh–Ptc (50%:80%; Fig. 4d and data not shown). Thus, as little as an approximately twofold change in the ratio of Hh–Ptc:Ptc^{Δloop2} seems to be sufficient to distinguish between on and off states of the pathway. We note that the total amount of Ptc protein in these clones falls within a range of ~130–360% relative to the basal level of endogenous Ptc, well below the level of peak Ptc expression induced by Hh in anterior cells abutting the posterior compartment (~700%). The range in the amount of Ptc corresponds to the moderately elevated levels of endogenous Ptc that are present normally in cells responding to intermediate levels of Hh signal. Hence, relatively small (on the scale of a few fold) differences in the ratio of liganded:unliganded Ptc may be sufficient to control the full

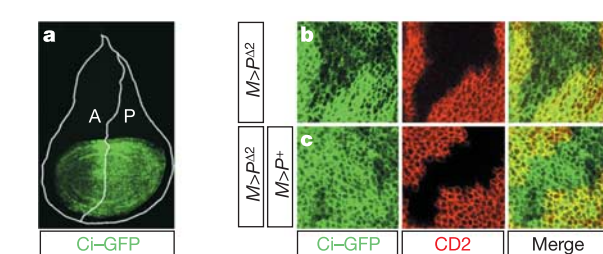


Figure 3 Inhibition of Hh transduction by Ptc^{Δloop2} is blocked by co-expression of Ptc⁺ and Hh. **a**, *CiΔZnGFP* (Ci–GFP, green) expressed throughout the wing pouch under the control of *nubbin-Gal4* is destabilized in anterior compartment cells away from the compartment boundary. **b**, $M > P^{\Delta 2}$ clone (marked black by the absence of CD2 (red)) in the posterior compartment where endogenous Ptc is not expressed: Ci–GFP is destabilized. **c**, $M > P^{\Delta 2}$ clone, similarly marked, in the posterior compartment of a wing disc expressing Ptc⁺ from a $M > P^+$ transgene³: Ci–GFP is not destabilized.

range of responses to the Hh gradient under physiologically relevant conditions.

We note that our results differ from those of Taipale *et al.*¹⁷, who found that overexpressing mPtc⁺ and mPtc^{Δloop2} in a 3:1 ratio in mammalian tissue culture cells exposed to a soluble amino-terminal form of Sonic Hh (Shh–N) failed to activate target gene expression. However, the actual ratio of liganded to unliganded forms of mPtc

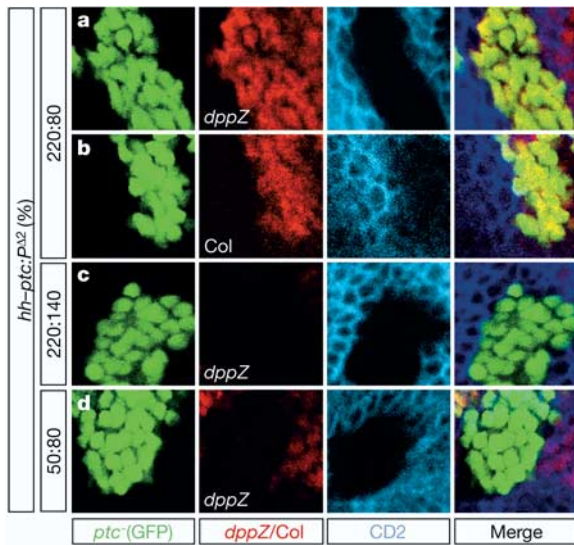


Figure 4 Estimating the change in the ratio of liganded to unliganded Ptc necessary to distinguish on and off states of the Hh transduction pathway. **a–d**, Anterior compartment clones of *ptc*[−] cells (marked by the presence of GFP (green)) expressing given combinations of *L* > *P*^{Δ2} or *M* > *P*^{Δ2} and *L* > *hh-ptc* or *H* > *hh-ptc* (marked black by the absence of CD2 staining (blue)) were monitored for the expression of *dpp-lacZ* (**a, c, d**; red) or Col (**b**; red). **a, b**, *dpp-lacZ* (**a**) and Col (**b**) are both strongly expressed in *ptc*[−] cells co-expressing high levels of Hh-Ptc (*H* > *hh-ptc*, ~220%; *n* = 7; s.d. = 44%) and low levels of Ptc^{Δloop2} (*L* > *P*^{Δ2}, ~80%). **c**, Increasing the level of Ptc^{Δloop2} from low (~80%) to medium (*M* > *P*^{Δ2}, ~140%) while holding the high level of Hh-Ptc (~220%) constant was sufficient to inhibit the pathway in *ptc*[−] cells, as shown by the absence of *dpp-lacZ* expression. **d**, Decreasing the level of Hh-Ptc from high (~220%) to low (*L* > *hh-ptc*, ~50%; *n* = 8; s.d. = 11%) while holding the low level (~80%) of Ptc^{Δloop2} expression constant also resulted in inhibition of the pathway in *ptc*[−] cells (*dpp-lacZ* expression is no longer detected).

obtained in this experiment was not determined; if as little as one-fifth of the available mPtc⁺ remained unliganded, the ratio of liganded to unliganded forms of mPtc protein would be 3:2, which our present results (Fig. 4c) suggest may be inadequate to activate Hh target genes under normal conditions. It is also unclear whether the ratio of liganded to unliganded Ptc necessary to activate the pathway will be the same in all biological contexts, or when the total amount of Ptc protein is well above peak physiological levels, as is likely in the tissue culture experiments.

How might the ratio of liganded to unliganded Ptc determine the level of Smo activity? Unliganded Ptc might function catalytically to inhibit Smo^{16,17,22–24}, and the presence of liganded Ptc would titrate or counterbalance the catalytic activity of unliganded Ptc. For example, unliganded Ptc might function as a transporter to drive the cytosolic accumulation of a Smo antagonist¹⁷, and liganded Ptc might allow some antagonist to translocate in the opposite direction. Alternatively, liganded and unliganded Ptc might function stoichiometrically by competing for access to Smo (or a Smo effector) and exerting opposing effects on its activity^{12,16}. Another possibility is that Ptc exists as a higher-order multimer (for example, as a trimer), as proposed for the structurally related AcrB transporter²⁸. In this model, binding of Hh to one monomer in any given multimer might block the ability of that multimer to function catalytically to inhibit Smo, or interfere with the activity of the remaining monomers, for example, by targeting the multimer for degradation.

A conserved feature of Hh signalling is that, during normal development, high levels of Hh signalling upregulate *ptc* transcription¹⁶. This generates a sharp peak of Ptc accumulation^{2,7,8,19,20} (Fig. 1c) that has an essential role in sequestering Hh and limiting its spread into responsive tissue^{3,12}. A ratiometric mechanism that

involves titration of the inhibitory action of unliganded Ptc by liganded Ptc might allow cells close to the source of Hh to upregulate sufficiently high levels of Ptc in order to impede effectively the movement of Hh without suppressing the response of these same cells to Hh by the inhibitory action of any residual unbound Ptc. Our present results suggest that as the gradient of Hh declines from peak levels to nil, the level of Smo activity, and hence the cell's perception of ambient Hh, will depend on the ratio of liganded to unliganded Ptc, which in turn will depend on the absolute amount of Ptc. In regions away from the Hh source, where Ptc is upregulated weakly or not at all, titration of unbound Ptc by Hh-bound Ptc might also be necessary for cells to detect and respond appropriately to low levels of ambient Hh. □

Methods

Immunostaining

Immunostaining was performed using standard techniques with antibodies against Ptc (I. Guerrero⁸), CD2 (Serotec), βgal (Cappel) and Col (A. Vincent²¹).

Mutations and transgenes used

Unless otherwise stated, all mutations and transgenes have been described previously, or were derived by combining well-defined components of previously described transgenes (see refs 3, 6, 12, 29). The *H* > *P*^{Δ2} transgene was generated *in vivo* by Flp-mediated recombination of a *Tuba1* > *CD2,y⁺* > *ptc*^{Δloop2} transgene containing the ubiquitously expressed *Tubulinα1* (*Tuba1*) promoter: recombination excises the > *CD2,y⁺* > Flp-out cassette to create clones of *Tuba1* > *ptc*^{Δloop2} (*H* > *P*^{Δ2}) cells, which express Ptc^{Δloop2} instead of the reporter protein CD2. The *M* > *P*^{Δ2} and *L* > *P*^{Δ2} transgenes were generated similarly. The *M* > *P*^{Δ2} transgene differs from *H* > *P*^{Δ2} in using the *hsp70* 3' untranslated region (UTR), which is ~two–threefold less stable than the *Tuba1* 3' UTR used in *H* > *P*^{Δ2} (ref. 29); the *L* > *P*^{Δ2} transgene differs from *M* > *P*^{Δ2} in using the *ribosomal protein 49* (*rp49*) promoter (which is ~twofold weaker than the *Tuba1* promoter)²⁹ (Fig. 1b). Flp-out and mitotic recombination clones were generated using standard techniques (for example, as in refs 3, 6, 30); multiple heat shocks were given during the first instar to generate clones within clones (as in Figs 2d and 4).

The *ciΔZnGFP* sequence encodes an inactive form of Ci, lacking the third and fourth zinc fingers (deleted between Pro 498 and Ser 594), fused to GFP at the carboxy terminus, and is expressed under Gal4/UAS control.

The Hh-Ptc protein is composed of the N-terminal signalling portion of Hh (amino acids 1–256) joined via three haemagglutinin (HA) tags to the juxtamembrane and transmembrane domains of the receptor Sevenless (Sev), joined to the N terminus of full-length Ptc (Fig. 1a; HA-Sev and Sev-Ptc joins are, respectively, ISSHHVH-LST and LVRRK-rsgsls-RDLSLP; with linker peptide in lower case). Hh-Ptc behaves as a constitutively liganded form of Ptc by the following criteria (data not shown). First, overexpression of Hh-Ptc does not rescue the absence of endogenous Ptc (expected, as Hh-Ptc should be self-inactivated; genotype: *y w hsp70-flp UAS-GFPnls; FRT42D ptc^{Δloop2} / dpp-lacZ¹⁰⁶²⁸ FRT42D Tuba1-Gal80; UAS-hh-ptc/Tuba1-Gal4*). Second, overexpression of Hh-Ptc ectopically activates the Hh transduction pathway in otherwise wild-type cells in the wing primordium (expected if Hh-Ptc titrates the inhibitory action of endogenous, unliganded Ptc; genotype: *y w hsp70-flp UAS-GFPnls; dpp-lacZ¹⁰⁶²⁸ / +; UAS-hh-ptc/Tuba1 > CD2,y⁺ > Gal4*). In contrast, clones similarly overexpressing a deleted form of Hh-Ptc lacking the Hh signalling domain do not activate the Hh transduction pathway (expected if the Hh signalling domain is responsible for inactivating Ptc within the Hh-Ptc chimera). Third, clones overexpressing Hh-Ptc constitutively activate the Hh pathway in a strictly cell autonomous fashion (expected if the Hh domain is sequestered by the Ptc protein to which it is fused and not available to bind Ptc on neighbouring cells; genotype as in the second criterion). In contrast, clones similarly overexpressing a deleted form of Hh-Ptc that lacks the second extracellular loop (Hh-Ptc^{Δloop2}) activate the Hh pathway in neighbouring cells (expected, as the attached Hh signalling domain should be available to signal to neighbouring cells because it cannot be bound by Ptc^{Δloop2} in the Hh-Ptc^{Δloop2} chimera).

Expression level quantification

For each posterior compartment clone of *L* > *P*^{Δ2}, *M* > *P*^{Δ2} and *H* > *P*^{Δ2} cells, the level of Ptc^{Δloop2} expression relative to the basal level of endogenous Ptc expression in the anterior compartment was determined by averaging the intensity of five samples of 35 pixels each within (1) the clone; (2) the anterior compartment far from the compartment boundary; and (3) portions of the posterior compartment not containing the clone. The resulting levels of staining in the clone (1) and in the anterior compartment (2) were normalized by subtracting the background level of staining (3); the level of staining in the clone is expressed as the per cent of the level of basal staining in the anterior compartment. All measurements were made within the wing primordium and staining intensity was quantified using ImageJ 1.25 (Rasband, W. S., <http://rsb.info.nih.gov/ij/>). To validate the method, we quantified the levels of Ptc staining in 2 × *ptc*⁺ clones relative to surrounding 1 × *ptc*⁺ (*FRT42D ptc^{Δloop2} / FRT42D arm-lacZ*) tissue (normalizing the level of staining by subtracting the level of staining detected in the 0 × *ptc*⁺ 'twin-spot' clones); the level of Ptc expression in 1 × *ptc*⁺ cells was 51% (*n* = 5; s.d. = 15%) that of neighbouring 2 × *ptc*⁺ cells. We also quantified Ptc staining in anterior compartment clones of *M* > *P*^{Δ2} cells. The measurement obtained, 210% (*n* = 4; s.d. = 22%), corresponds approximately to the

sum of Ptc staining in $M > P^{42}$ posterior compartment clones and the basal level of Ptc staining in anterior compartment cells (Fig. 1 legend), validating our use of posterior compartment clones to quantify Ptc ^{Δ loop2} expression in sibling clones in the anterior compartment.

Genotypes

We used the following *Drosophila* genotypes. Figures 1 and 2a–c: $y w hsp70\text{-}flp; +$ (or $dpp\text{-}lacZ^{10628}/+$; $Tub\alpha 1$ (or $rp49$) $> CD2, y^+ > ptc^{\Delta loop 2}\text{-}Tub\alpha 1$ 3' UTR (or $hsp70$ 3' UTR)/+). Figure 2d: $y w hsp70\text{-}flp$ UAS–GFPnls; FRT42D $ptc^{IIW}/dpp\text{-}lacZ^{10628}$ FRT42D $Tub\alpha 1\text{-}Gal80$; $rp49 > CD2, y^+ > ptc^{\Delta loop 2}\text{-}hsp70$ 3' UTR/ $Tub\alpha 1\text{-}Gal4$. Figure 2e–g: $y w hsp70\text{-}flp$; FRT42D ptc^{IIW} $sha/FRT42$ P($w^+ ptc^+$) P($hsp70\text{-}CD2, y^+$); $rp49 > ptc^{\Delta loop 2}\text{-}hsp70$ 3' UTR/+. P($w^+ ptc^+$), a gift from J. Hooper, contains a ~20-kilobase genomic fragment from the *ptc* locus that confers partial rescuing activity. Figure 3: $y w hsp70\text{-}flp$; $nub\text{-}Gal4$ UAS– $\Delta ZnfGFP/+$; $Tub\alpha 1 > CD2, y^+ > ptc^{\Delta loop 2}\text{-}hsp70$ 3' UTR/+ (or $Tub\alpha 1 > ptc^+$ $\text{-}hsp70$ 3' UTR). Figure 4: $y w hsp70\text{-}flp$ UAS–GFPnls; $rp49$ (or $Tub\alpha 1$) $> CD2, y^+ > ptc^{\Delta loop 2}\text{-}hsp70$ 3' UTR FRT42D $ptc^{IIW}/dpp\text{-}lacZ^{10628}$ FRT42D $Tub\alpha 1\text{-}Gal80$; $Tub\alpha 1 > CD2, y^+ > hh\text{-}ptc\text{-}Tub\alpha 1$ 3' UTR (or $rp49 > CD2, y^+ > hh\text{-}ptc\text{-}hsp70$ 3' UTR)/ $Tub\alpha 1\text{-}Gal4$.

Received 12 February; accepted 9 July 2004; doi:10.1038/nature02835.

Published online 8 August 2004.

- Gurdon, J. B. & Bourillot, P. Y. Morphogen gradient interpretation. *Nature* **413**, 797–803 (2001).
- Ingham, P. W., Taylor, A. M. & Nakano, Y. Role of the *Drosophila* patched gene in positional signalling. *Nature* **353**, 184–187 (1991).
- Chen, Y. & Struhl, G. Dual roles for patched in sequestering and transducing Hedgehog. *Cell* **87**, 553–563 (1996).
- Stone, D. M. *et al.* The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**, 129–134 (1996).
- Marigo, V., Davey, R. A., Zuo, Y., Cunningham, J. M. & Tabin, C. J. Biochemical evidence that patched is the Hedgehog receptor. *Nature* **384**, 176–179 (1996).
- Basler, K. & Struhl, G. Compartment boundaries and the control of *Drosophila* limb pattern by hedgehog protein. *Nature* **368**, 208–214 (1994).
- Tabata, T. & Kornberg, T. B. Hedgehog is a signaling protein with a key role in patterning *Drosophila* imaginal discs. *Cell* **76**, 89–102 (1994).
- Capdevila, J., Estrada, M. P., Sanchez-Herrero, E. & Guerrero, I. The *Drosophila* segment polarity gene patched interacts with decapentaplegic in wing development. *EMBO J.* **13**, 71–82 (1994).
- Struhl, G., Barbash, D. A. & Lawrence, P. A. Hedgehog organizes the pattern and polarity of epidermal cells in the *Drosophila* abdomen. *Development* **124**, 2143–2154 (1997).
- Mullor, J. L., Calleja, M., Capdevila, J. & Guerrero, I. Hedgehog activity, independent of decapentaplegic, participates in wing disc patterning. *Development* **124**, 1227–1237 (1997).
- Strigini, M. & Cohen, S. M. A Hedgehog activity gradient contributes to AP axial patterning of the *Drosophila* wing. *Development* **124**, 4697–4705 (1997).
- Briscoe, J., Chen, Y., Jessell, T. M. & Struhl, G. A hedgehog-insensitive form of patched provides evidence for direct long-range morphogen activity of sonic hedgehog in the neural tube. *Mol. Cell* **7**, 1279–1291 (2001).
- Hooper, J. E. Distinct pathways for autocrine and paracrine Wingless signalling in *Drosophila* embryos. *Nature* **372**, 461–464 (1994).
- van den Heuvel, M. & Ingham, P. W. *smoothed* encodes a receptor-like serpentine protein required for hedgehog signalling. *Nature* **382**, 547–551 (1996).
- Alcedo, J., Ayzenzon, M., Von Ohlen, T., Noll, M. & Hooper, J. E. The *Drosophila* *smoothed* gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. *Cell* **86**, 221–232 (1996).
- Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001).
- Taipale, J., Cooper, M. K., Maiti, T. & Beachy, P. A. Patched acts catalytically to suppress the activity of *Smoothed*. *Nature* **418**, 892–897 (2002).
- Lawrence, P. A. & Struhl, G. Morphogens, compartments, and pattern: lessons from *Drosophila*? *Cell* **85**, 951–961 (1996).
- Nakano, Y. *et al.* A protein with several possible membrane-spanning domains encoded by the *Drosophila* segment polarity gene *patched*. *Nature* **341**, 508–513 (1989).
- Hooper, J. E. & Scott, M. P. The *Drosophila* patched gene encodes a putative membrane protein required for segmental patterning. *Cell* **59**, 751–765 (1989).
- Vervoort, M., Crozatier, M., Valle, D. & Vincent, A. The COE transcription factor Collier is a mediator of short-range Hedgehog-induced patterning of the *Drosophila* wing. *Curr. Biol.* **9**, 632–639 (1999).
- Alcedo, J., Zou, Y. & Noll, M. Posttranscriptional regulation of *smoothed* is part of a self-correcting mechanism in the Hedgehog signaling system. *Mol. Cell* **6**, 457–465 (2000).
- Denef, N., Neuberger, D., Perez, L. & Cohen, S. M. Hedgehog induces opposite changes in turnover and subcellular localization of patched and *smoothed*. *Cell* **102**, 521–531 (2000).
- Ingham, P. W. *et al.* Patched represses the Hedgehog signalling pathway by promoting modification of the *Smoothed* protein. *Curr. Biol.* **10**, 1315–1318 (2000).
- Hepker, J., Wang, Q. T., Motzny, C. K., Holmgren, R. & Orenic, T. V. *Drosophila* cubitus interruptus forms a negative feedback loop with patched and regulates expression of Hedgehog target genes. *Development* **124**, 549–558 (1997).
- Aza-Blanc, P. & Kornberg, T. B. Ci: a complex transducer of the hedgehog signal. *Trends Genet.* **15**, 458–462 (1999).
- Methot, N. & Basler, K. Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. *Cell* **96**, 819–831 (1999).
- Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587–593 (2002).
- Greenwood, S. & Struhl, G. Different levels of Ras activity can specify distinct transcriptional and morphological consequences in early *Drosophila* embryos. *Development* **124**, 4879–4886 (1997).
- Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).

Acknowledgements We thank A. Adachi for generating transgenic flies; X.-J. Qiu for technical assistance; and R. Axel, E. Gouaux, I. Greenwald, T. Jessell, L. Johnston, R. Mann and A. Tomlinson for discussion and advice on the manuscript. A.C. is a Postdoctoral Associate and G.S. is an Investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.S. (gs20@columbia.edu).

Netrin-1 controls colorectal tumorigenesis by regulating apoptosis

Laetitia Mazelin¹, Agnès Bernet^{1*}, Christelle Bonod-Bidaud^{1*}, Laurent Pays¹, Ségolène Arnaud¹, Christian Gespach², Dale E. Bredesen³, Jean-Yves Scoazec⁴ & Patrick Mehlen^{1,5}

¹Apoptosis/Differentiation Laboratory—Equipe labellisée ‘La Ligue’—Molecular and Cellular Genetic Center, CNRS UMR 5534, University of Lyon, 69622 Villeurbanne, France

²INSERM U482, Signal Transduction and Cellular Functions in Diabetes and Digestive Cancers, Hospital Saint-Antoine, 75571 Paris, France

³The Buck Institute for Age Research, Novato, California 94945, USA

⁴INSERM U45 and ANIPATH, IFR62, Faculté Laennec, 69437 Lyon, France

⁵Centre Leon Bérard, 69373 Lyon, France

* These authors contributed equally to this work.

The expression of the protein DCC (deleted in colorectal cancer) is lost or markedly reduced in numerous cancers and in the majority of colorectal cancers due to loss of heterozygosity in chromosome 18q, and has therefore been proposed to be a tumour suppressor¹. However, the rarity of mutations found in DCC, the lack of cancer predisposition of DCC mutant mice, and the presence of other tumour suppressor genes in 18q have raised doubts about the function of DCC as a tumour suppressor². Unlike classical tumour suppressors, DCC has been shown to induce apoptosis conditionally: by functioning as a dependence receptor, DCC induces apoptosis unless DCC is engaged by its ligand, netrin-1 (ref. 3). Here we show that inhibition of cell death by enforced expression of netrin-1 in mouse gastrointestinal tract leads to the spontaneous formation of hyperplastic and neoplastic lesions. Moreover, in the adenomatous polyposis coli mutant background associated with adenoma formation, enforced expression of netrin-1 engenders aggressive adenocarcinomatous malignancies. These data demonstrate that netrin-1 can promote intestinal tumour development, probably by regulating cell survival. Thus, a netrin-1 receptor or receptors function as conditional tumour suppressors.

Netrin-1, a diffusible laminin-related protein, is a bifunctional molecule. It has a major role during nervous system development in mediating chemo-attraction and chemo-repulsion of axons and neurons by interacting with its main receptor, DCC (refs 4, 5, 6). However, netrin-1 has also been described recently as a survival factor. Indeed, the netrin-1 receptors DCC and UNC5H (that is, UNC5H1, UNC5H2 and UNC5H3) belong to the so-called dependence receptor family^{3,7}. Such receptors, which also include RET (rearranged during transfection)⁸, β -integrins⁹, Patched¹⁰ and the p75 neurotrophin receptor (p75NTR)¹¹, share the functional property of inducing cell death when disengaged from their ligands but not when bound by their ligands. Such receptors thus create cellular states of dependence on their respective ligands^{12,13}.

This pattern of cell death induction (although to date described predominantly in cell culture experiments) has led to the proposal that these receptors' pro-apoptotic activity may constitute a regulatory mechanism that would eliminate cells growing in inappropriate settings in which the ligand is unavailable. This hypothesis is compatible with the fact that DCC was proposed in the early 1990's to be a tumour suppressor gene mainly associated with the later stages of intestinal tumour development^{1,14}. Here we considered the possibility that, unlike classical tumour suppressors, DCC may represent an alternative type of tumour suppressor, whose activity is conditional, by functioning as a dependence receptor.

In initial studies, we investigated whether or not the pro-apoptotic activity of DCC observed in the absence of netrin-1 may be a regulatory mechanism that supports the putative tumour suppressor function of DCC. We observed that the known ability of DCC to inhibit the hallmarks of cell transformation *in vitro*¹⁵ might be attributable to DCC pro-apoptotic activity because, in various cancer cell lines, although DCC expression constrained the anchorage-independent growth of these cells, both netrin-1 presence and caspase inhibition independently reversed this growth suppression by DCC (Supplementary Fig. 1; and data not shown).

Because netrin-1 inhibited both DCC's pro-apoptotic effect³ and its tumour suppressive activity *in vitro* (Supplementary Fig. 1), we next determined whether netrin-1 is indeed expressed in the gastrointestinal epithelium. Although netrin-1 was barely detectable by northern blot unless poly(A)⁺ messenger RNA was probed¹⁶, more sensitive methods such as polymerase chain reaction with reverse transcription (RT-PCR) or *in situ* hybridization allowed the detection of netrin-1 in murine gastrointestinal tracts (Fig. 1a–c). Moreover, the endogenous protein was detected by immunohistochemistry (Fig. 1d, e). As shown in Fig. 1, both netrin-1 mRNA and

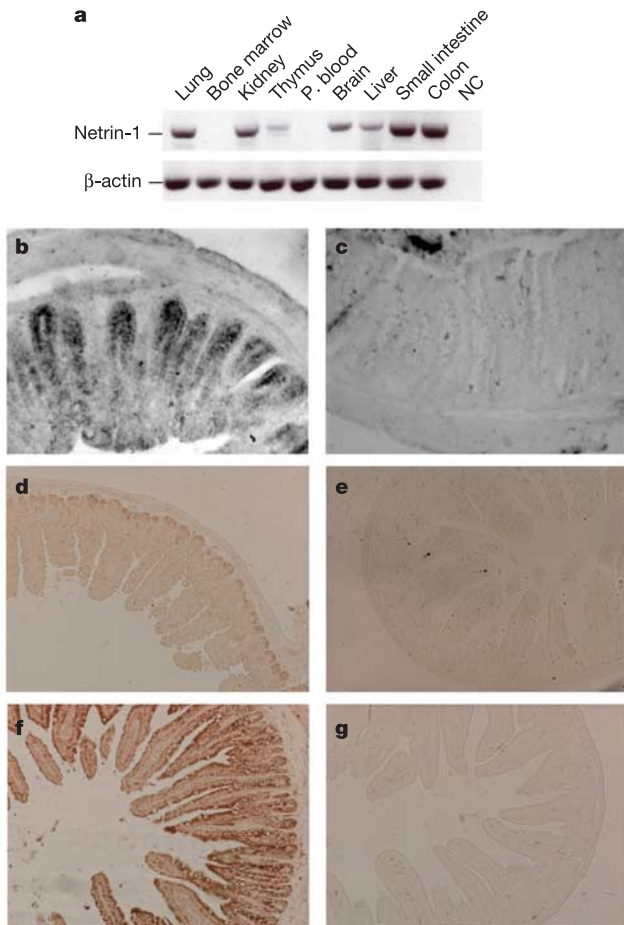


Figure 1 Netrin-1 is expressed in the mouse gastrointestinal tract. **a**, RT-PCR using specific mouse netrin-1 primers was performed using total RNA extracted from various tissues as indicated. NC, negative control. **b**, **c**, *In situ* hybridization was performed on 6-month-old mouse colons as described in the Supplementary Methods section using either an anti-sense netrin-1 probe (**b**) or control sense netrin-1 probe (**c**). Note the specific staining at the base of the colonic crypts. **d**, **e**, Immunohistochemistry on 6-month-old mouse small intestine using anti-netrin-1 antibody (**d**) or no primary antibody (**e**). **f**, **g**, Immunohistochemistry on 6-month-old mouse small intestine using anti-DCC antibody without (**f**) or with prior incubation with a competitive peptide (**g**). Original magnifications: **b**, **c**, $\times 120$; **d**–**g**, $\times 100$.

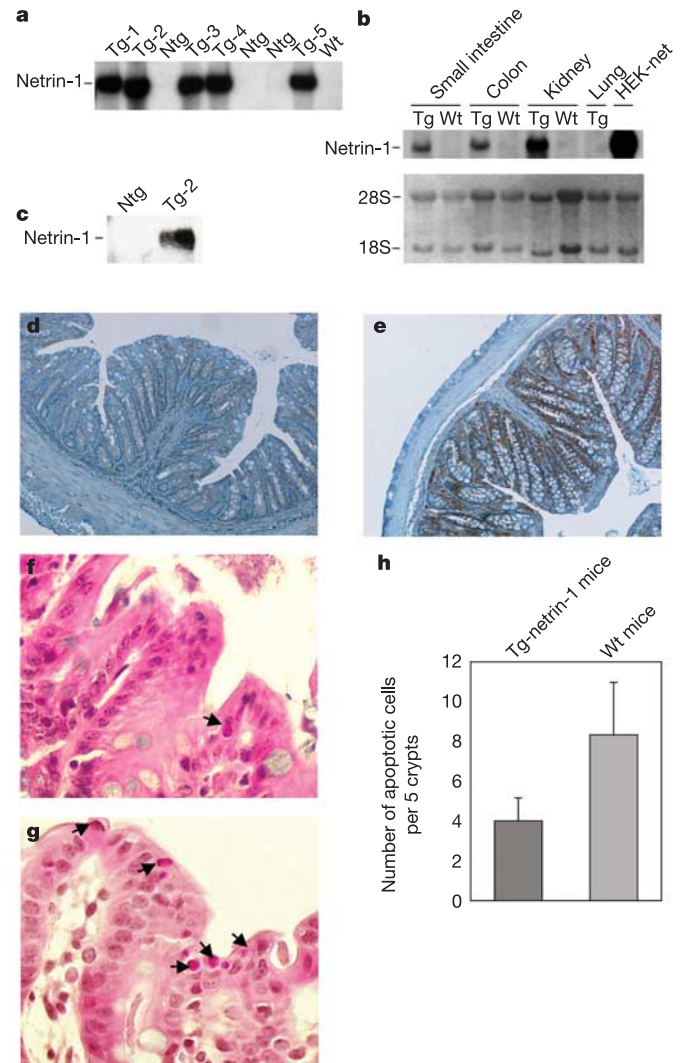


Figure 2 Netrin-1 overexpression inhibits intestinal cell apoptosis. **a**, Tg-netrin-1 mouse tail DNA was analysed by Southern blot after restriction by *EcoRI* and *XhoI* using netrin-1 probe. A fragment of 2.3 kilobase pairs is expected if the Fabp1-netrin-1 construct is inserted. Tg-1 to 5, transgenic lines; ntg, non-transgenic line; wt, parental line. **b**, Total RNA from various tissues from either parental mice or Tg-netrin-1 mice were analysed by northern blot using a chick netrin-1 probe. **c**, Immunoblot performed after immunoprecipitation of c-Myc-tagged netrin-1 from dissociated villi and crypts of the small intestine from either non-transgenic mice or Tg-netrin-1 mice. **d**, **e**, Immunohistochemistry of tagged netrin-1 on colon section from Tg-netrin-1 mice (**e**) or control mice (**d**). **f**, **g**, Tg-netrin-1 mice showed a decrease in intestinal cell death. Representative microscopic images from hematoxylin-eosin-saffron staining of 20-month-old Tg-netrin-1 mouse (**f**) or wild-type mouse (**g**) colon section showing apoptotic cell death. Arrows indicate cells with apoptotic morphology. Quantitative analysis of apoptotic cells in either wild-type or Tg-netrin-1 mice performed from the observations of 14 animals from each type (**h**). Standard deviations are indicated ($P < 0.0001$). Original magnifications: **d**, $\times 100$; **e**, $\times 120$; **f**, **g**, $\times 450$.

protein were detected, mainly in crypts of the small intestine and at the bases of the colonic epithelial crypts; DCC, in contrast, was found to be expressed all along the intestinal epithelium (Fig. 1f, g), as previously shown in human gut¹⁷. This restricted expression of netrin-1, together with its diffusible character, suggest that a netrin-1 gradient may exist, with higher concentrations of netrin-1 present at the proximal portion of the intestinal villus—known to be a site of intense cellular proliferation—than at the distal tip of the villus, a site of cell death and shedding.

To assess whether netrin-1 regulates DCC-induced apoptosis in the intestinal epithelium, we generated transgenic mice that over-expressed netrin-1 in the gut. A series of independent lines of transgenic mice were obtained that had inserted a transgene containing netrin-1 complementary DNA under the gut-specific Fabp1(4x at -132) promoter¹⁸ (Fig. 2a). Five of these lines (Tg-netrin-1/1 to 5) were analysed further (Fig. 2a). Figure 2b shows expression of the transgene in the gut—both small intestine and colon—but not in the lung. Moreover, ectopically expressed

tagged netrin-1 was detected by western blot in epithelial villi and crypts isolated from the small intestine of Tg-netrin-1 mice (Fig. 2c). Similarly, immunohistochemistry showed expression of the exogenous netrin-1 all along the intestinal epithelium (Fig. 2d, e). Together with foregut and hindgut expression, exogenous netrin-1 expression was also detected in the kidney (Fig. 2b). The five lines of Tg-netrin-1 mice studied developed and bred normally without any obvious gastrointestinal defects.

Cell death in the gut of these five lines of mice was then analysed in a series of 14 mice from either normal or Tg-netrin-1 adult mice. As shown in Fig. 2f–h, a 50% decrease of apoptotic cells was observed in the intestinal epithelia of Tg-netrin-1 mice ($P < 0.0001$), although no obvious change in the status of proliferation (assessed by BrdU incorporation) or differentiation (assessed by Goblet cell-specific staining) was observed (Supplementary Fig. 2). Thus, rather than inducing ectopic signalling for proliferation or differentiation, netrin-1 overexpression, in agreement with predictions from previous *in vitro* studies of dependence receptors^{3,12}, inhibited death signalling in the intestinal epithelium.

To determine whether this netrin-1-dependent decrease in cell death engenders tumour development, the intestinal tracts from a series of 23 normal and 30 Tg-netrin-1 20-month-old mice were analysed for colorectal tumours (Fig. 3). Although no colorectal adenocarcinoma-like malignancies were detected in either mouse population, 17% of the Tg-netrin-1 mice displayed at least one adenomatous lesion, located in the small intestine in 10% of animals and in the colon in 7% of animals, whereas such neoplasms were not detected in any of the control mice. Moreover, 43% of

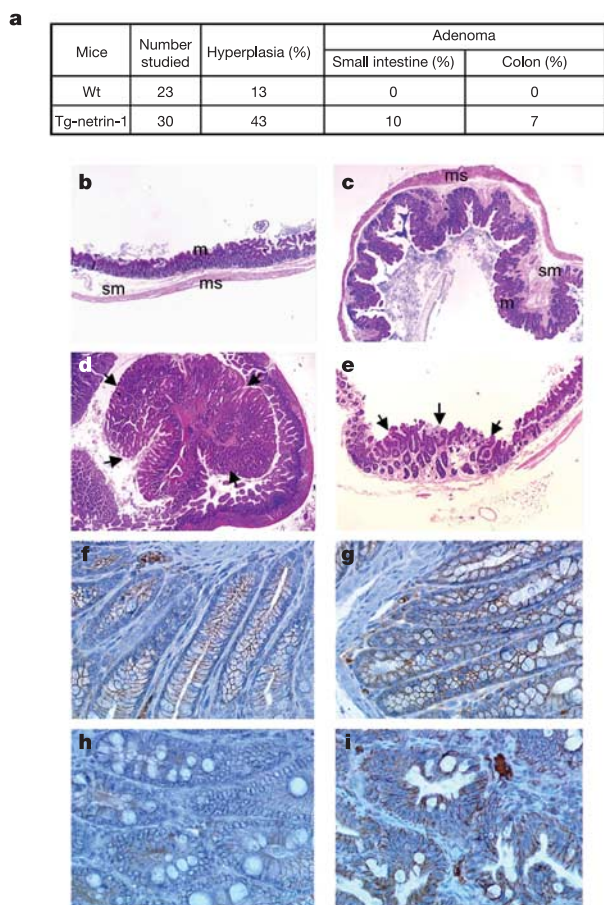


Figure 3 Netrin-1 overexpression enhances the early stage of tumour development. **a**, Table showing the number of mice studied and the proportion of hyperplasia and adenoma in the small intestine and colon. 5 lines of Tg-netrin-1 mice were analysed. **b, c, d, e**, Hematoxylin-eosin-saffron staining of the small intestine and colon on Tg-netrin-1 and wild type animals. Compared with the normal colonic mucosa shown in **b**, in 43% Tg-netrin-1 animals, diffuse hyperplasia of the colonic mucosa (**c**) was present; it was characterized by an increase in the height of the mucosal layer and by the formation of pseudo-villous structures. m, mucosa; sm, submucosa; ms, muscularis. In some animals, low-grade adenomas were present, either in the small intestine (**d**, low-grade polypoid adenoma, arrows) or in the colon (**e**, low-grade sessile adenoma, arrows). **f–i**, β -catenin immunohistochemical localization on normal colonic mucosa (**f**), Tg-netrin-1 colonic mucosa (**g**), hyperplasia of Tg-netrin-1 (**h**) or adenoma of Tg-netrin-1 (**i**). Original magnifications: **b–e**, $\times 60$; **f–i**, $\times 350$.

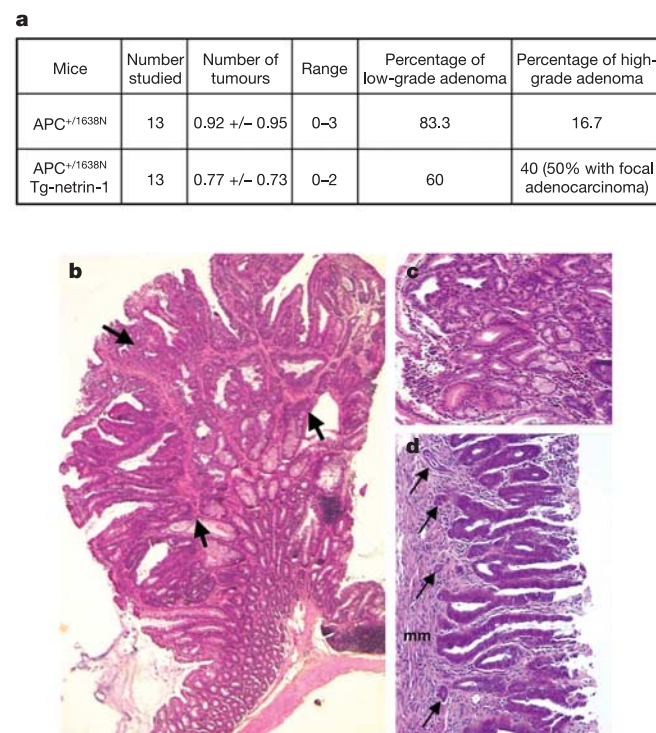


Figure 4 Netrin-1 overexpression enhances the adenoma to adenoma-carcinoma transition. **a**, Table showing the number of APC^{+/1638N} or APC^{+/1638N}/Tg-netrin-1 mice studied and the proportion of low-grade adenoma and high-grade adenoma. Two independent lines of Tg-netrin-1 mice were backcrossed into the APC^{+/1638N} background and analysed for tumour development. **b, c, d**, Hematoxylin-eosin-saffron staining of: polypoid high-grade adenoma (**b**, arrows); a high power view of a low-grade adenoma (**c**); and of focal invasion of the muscularis mucosae (mm) by small carcinomatous glands in the deep part of a mucosal adenocarcinomatous lesion (**d**). Original magnifications: **b**, $\times 100$; **c**, $\times 280$; **d**, $\times 300$.

Tg-netrin-1 mice showed focal or diffuse hyperplasia of the colon mucosa, characterized by an increase in the height of the mucosal layer and formation of pseudo-villous structures (Fig. 3c); in contrast, only 13% of control animals showed mucosal hyperplasia ($P < 0.001$). Interestingly, the Wnt–catenin signalling did not seem to be modulated in these hyperplastic or adenomatous lesions, because a typical localization of β -catenin at lateral cell–cell contacts was still observed in these structures (Fig. 3f–i).

Because DCC loss is observed with high frequency in late stage tumours, suggesting that the putative tumour suppressive role of DCC may modulate a predominantly late event in tumourigenesis¹⁹, we investigated the possibility that netrin-1 overexpression may affect tumour progression. Tg-netrin-1 mice were backcrossed onto an adenomatous polyposis coli (APC)^{+/1638N} genetic background²⁰. APC mutations exist in the vast majority of human colorectal tumours and in mice, specific APC mutations trigger the formation of low-grade adenomas. The number of neoplastic lesions per control mouse in this APC^{+/1638N} genetic background was 0.92, with a predominance of low-grade adenomas. A striking difference was observed, however, in Tg-netrin-1 APC^{+/1638N} mice (Fig. 4). Although these mice displayed a similar overall frequency of neoplastic lesions to the APC^{+/1638N} mice, they manifested a significantly higher incidence of high-grade adenoma (40% versus 16.7%, $P < 0.001$). Furthermore, 50% of high-grade adenomas in the Tg-netrin-1 APC^{+/1638N} mice contained foci of epithelial lesions that were classified histologically as adenocarcinomas, either intra-epithelial or invasive, with focal invasion of the muscularis mucosa (Fig. 4d). Thus, netrin-1-mediated apoptosis inhibition is associated with an enhancement of tumour progression in the presence of an APC mutation.

In summary, netrin-1 is constitutively expressed in colorectal tissue, with preferential expression at the bases of intestinal crypts. Ectopic expression of netrin-1 reduced intestinal cell death, whereas the netrin-1 mutant (hypomorphic) newborn mice displayed an increased cell death (0.12 ± 0.13 versus 1.2 ± 0.23 apoptotic cells per crypt in control versus netrin-1^{-/-} mice, respectively; see also Supplementary Fig. 3) in both the foregut and hindgut, albeit to a lesser extent in the large intestine (5.7-fold increase) than in the small intestine (13.2-fold increase). Thus, netrin-1 level is crucial to regulate intestinal cell death. However, it is noteworthy that, even though netrin-1 overexpression was shown to decrease intestinal cell death induction, global disorganization of the colorectal epithelium of these animals was not observed. Thus, netrin-1 regulation of cell death is unlikely to be responsible for a general homeostatic regulatory process that would balance colorectal epithelial cell proliferation and differentiation. Rather, one may speculate that the apparent proximal-to-distal netrin-1 gradient may function as a regulatory system that limits the lifespan of cells that undergo: (1) multiple proliferative steps within the crypt, and (2) repeated mechanical and chemical insults originating from the intestinal lumen, two conditions that may increase the risk of cell damage and resultant aberrant behaviour. Consequently, the cell death induction because of an absence of superficial netrin-1 expression, together with the mechanical detachment of cells in the lumen, may represent key factors for limiting the initiation of malignant transformation.

The current results provide a potential clue to the effect of the observed loss of DCC expression in the majority of malignant colorectal tumours. Loss of DCC expression would be predicted to induce a loss of cellular dependence on high concentrations of netrin-1 for survival. In this regard, it is of interest that netrin-1 dependence receptors other than DCC (that is, UNC5H1, UNC5H2 and UNC5H3) have recently been shown to be expressed in gut (ref. 21; and Supplementary Fig. 4) and decreased in expression in human colorectal cancers²¹. Moreover, the UNC5H2 gene was shown to be a direct transcriptional target of the tumour suppressor p53, and UNC5H2 was shown to mediate p53-induced cell death²².

Thus, both DCC and UNC5H loss of expression may contribute to the acquisition of malignant properties. It is thus intriguing to note that netrin-1 seems to have an effect at two different periods of this multi-step carcinogenesis process—that is, at an early stage (in light of the development of hyperplasia and adenoma) and at a later stage (in light of the shift towards adenocarcinoma in the APC-mutated background). One explanation may be that DCC and UNC5H receptors may be differentially important at these two periods, which would be concordant with a distinct intestinal localization of these receptors (Fig. 1f; and Supplementary Fig. 4). Therefore, the pro-apoptotic activities of these receptors may control colorectal tumourigenesis at two different steps of the process, and it may thus create a selective cellular survival advantage either to gain expression of netrin-1 (as observed in only a restricted number of colorectal tumours; see Supplementary Fig. 5) or to lose expression of these receptors. In the latter case, this may occur either early in neoplastic transformation (increasing the likelihood of adenoma development) or later in the process (increasing the likelihood of the neoplasms' undergoing the adenoma/adenocarcinoma transition). Consequently, DCC and UNC5H may be considered as tumour suppressors. However, they clearly differ from classical tumour suppressors, such as retinoblastoma (Rb), that constitutively inhibit tumour formation because of their intrinsic roles in cell cycle or other related properties. In contrast, dependence receptors may function as conditional tumour suppressors, because they acquire their suppressive activity only in the context of cell growth outside the region of high ligand availability, whereas potentially accelerating tumour development or progression within regions of inadequate high ligand concentration. □

Methods

Materials and Methods are presented in detail in Supplementary Information. Briefly, the construct used for transgenesis was obtained by PCR amplification of the Fabp1(4x at -132) promoter cassette and its insertion together with the full-length chick netrin-1 cDNA into pcDNA 3.1. A linear and purified DNA fragment of the pcDNA–Fabp1–netrin-1 construct was then microinjected into the male fertilized pronucleus obtained from C57BL/6 mice and mice founders were obtained from the SEAT transgenesis center (CNRS, Villejuif, France). Germline transmission and genotyping were detected by Southern-blot analysis of tail genomic DNA. APC^{+/1638N} mice were obtained from R. Fodde. *In situ* hybridization, RT–PCR, immunohistochemistry, immunoprecipitations and western blots were performed as described in Supplementary Information. Tumour analysis and apoptosis scoring were performed from hematoxylin–eosin–safran stained sections.

Received 22 April; accepted 25 June 2004; doi:10.1038/nature02788.

1. Fearon, E. R. *et al.* Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* **247**, 49–56 (1990).
2. Fazeli, A. *et al.* Phenotype of mice lacking functional Deleted in colorectal cancer (Dcc) gene. *Nature* **386**, 796–804 (1997).
3. Mehlen, P. *et al.* The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis. *Nature* **395**, 801–804 (1998).
4. Serafini, T. *et al.* Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**, 1001–1014 (1996).
5. Keino-Masu, K. *et al.* Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. *Cell* **87**, 175–185 (1996).
6. Forcet, C. *et al.* Netrin-1-mediated axon outgrowth requires deleted in colorectal cancer-dependent MAPK activation. *Nature* **417**, 443–447 (2002).
7. Llamas, F., Causeret, F., Bloch-Gallego, E. & Mehlen, P. Netrin-1 acts as a survival factor via its receptors UNC5H and DCC. *EMBO J.* **20**, 2715–2722 (2001).
8. Bordeaux, M. C. *et al.* The RET proto-oncogene induces apoptosis: a novel mechanism for Hirschsprung disease. *EMBO J.* **19**, 4056–4063 (2000).
9. Stupack, D. G., Puente, X. S., Boutsaboulay, S., Storgard, C. M. & Cheresch, D. A. Apoptosis of adherent cells by recruitment of caspase-8 to unligated integrins. *J. Cell Biol.* **155**, 459–470 (2001).
10. Thibert, C. *et al.* Inhibition of neuroepithelial patched-induced apoptosis by sonic hedgehog. *Science* **301**, 843–846 (2003).
11. Rabizadeh, S. *et al.* Induction of apoptosis by the low-affinity NGF receptor. *Science* **261**, 345–348 (1993).
12. Mehlen, P. & Bredesen, D. E. The dependence receptor hypothesis. *Apoptosis* **9**, 37–49 (2004).
13. Bredesen, D. E. *et al.* p75NTR and the concept of cellular dependence: seeing how the other half die. *Cell Death Differ.* **5**, 365–371 (1998).
14. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
15. Klingelhutz, A. J., Smith, P. P., Garrett, L. R. & McDougall, J. K. Alteration of the DCC tumor-suppressor gene in tumorigenic HPV-18 immortalized human keratinocytes transformed by nitrosomethylurea. *Oncogene* **8**, 95–99 (1993).
16. Meyerhardt, J. A. *et al.* Netrin-1: interaction with deleted in colorectal cancer (DCC) and alterations in brain tumors and neuroblastomas. *Cell Growth Differ.* **10**, 35–42 (1999).

17. Hsu, Y. H., Shaw, C. K. & Chuong, C. M. Immunohistochemical localization of deleted-in-colon-cancer (DCC) protein in human epithelial, neural, and neuro-endocrine tissues in paraffin sections with antigen retrieval. *Kaohsiung J. Med. Sci.* **17**, 351–357 (2001).
18. Simon, T. C., Cho, A., Tso, P. & Gordon, J. I. Suppressor and activator functions mediated by a repeated heptad sequence in the liver fatty acid-binding protein gene (Fabpl). Effects on renal, small intestinal, and colonic epithelial cell gene expression in transgenic mice. *J. Biol. Chem.* **272**, 10652–10663 (1997).
19. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
20. Fodde, R. *et al.* A targeted chain-termination mutation in the mouse Apc gene results in multiple intestinal tumors. *Proc. Natl Acad. Sci. USA* **91**, 8969–8973 (1994).
21. Thiebaut, K. *et al.* The netrin-1 receptors UNC5H are putative tumor suppressors controlling cell death commitment. *Proc. Natl Acad. Sci. USA* **100**, 4173–4178 (2003).
22. Tanikawa, C., Matsuda, K., Fukuda, S., Nakamura, Y. & Arakawa, H. p53RDL1 regulates p53-dependent apoptosis. *Nature Cell Biol.* **5**, 216–223 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We wish to thank N. Gadot, C. Guix and G. Perret for excellent technical assistance. We also thank R. Fodde and S. Robine for advice and materials. This work was supported by the Ligue Contre le Cancer (P.M.), the Schlumberger Fondation (P.M.), the NIH (to P.M. and D.E.B.) and the Region Rhone-Alpes (to P.M. and J.Y.S.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.M. (mehlen@univ-lyon1.fr).

A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response

Michael B. Reed¹, Pilar Domenech¹, Claudia Manca², Hua Su^{1*}, Amy K. Barczak¹, Barry N. Kreiswirth³, Gilla Kaplan² & Clifton E. Barry III¹

¹Tuberculosis Research Section, NIAID, National Institutes of Health, 12441 Parklawn Drive, Rockville, Maryland 20852, USA

²Laboratory of Mycobacterial Immunity and Pathogenesis, ³Tuberculosis Center, Public Health Research Institute, 225 Warren Street, Newark, New Jersey 07103-3535, USA

* Present address: Laboratory of Human Bacterial Pathogenesis, NIAID, NIH, 903 South 4th Street, Hamilton, Montana 59840, USA

Fifty million new infections with *Mycobacterium tuberculosis* occur annually, claiming 2–3 million lives from tuberculosis worldwide¹. Despite the apparent lack of significant genetic heterogeneity between strains of *M. tuberculosis*^{2,3}, there is mounting evidence that considerable heterogeneity exists in molecules important in disease pathogenesis. These differences may manifest in the ability of some isolates to modify the host cellular immune response, thereby contributing to the observed diversity of clinical outcomes^{4–7}. Here we describe the identification and functional relevance of a highly biologically active lipid species—a polyketide synthase-derived phenolic glycolipid (PGL) produced by a subset of *M. tuberculosis* isolates belonging to the W-Beijing family⁸ that show ‘hyperlethality’ in murine disease models. Disruption of PGL synthesis results in loss of this hypervirulent phenotype without significantly affecting bacterial load during disease. Loss of PGL was found to correlate with an increase in the release of the pro-inflammatory cytokines tumour-necrosis factor- α and interleukins 6 and 12 *in vitro*. Furthermore, the overproduction of PGL by *M. tuberculosis* or the addition of purified PGL to monocyte-derived macrophages was found to inhibit the release of these pro-inflammatory mediators in a dose-dependent manner.

Phenotypically distinct isolates of *M. tuberculosis* are able to differentially manipulate the delicate balance that exists between

the pathogen and the human host^{4–6}. For example, a tuberculosis outbreak in a rural community of the United States caused considerable anxiety due to the large number of diagnostic skin test conversions in casual contacts of the index case. The *M. tuberculosis* strain isolated, CDC1551, was initially reported to be hypervirulent based on this fact. Subsequently, however, this strain was not found to be associated with higher than normal rates of active tuberculosis⁷. Consistent with the high rate of delayed-type hypersensitivity reaction in humans, infection of mice with CDC1551 results in an augmented immediate pro-inflammatory cytokine response, coupled with longer survival of infected animals^{4,5}. Strain HN878, in contrast, was isolated from an outbreak in the Texas prison system and has been responsible for at least three clusters of active disease². When this strain was used to infect mice, a hypervirulent phenotype resulted wherein mice failed to induce a strong TH₁-type T-cell and cytokine response early in infection and died much more rapidly^{5,6}. It has also been demonstrated that the robust cytokine-inducing activity of CDC1551 in the mouse lung is reproducible *in vitro* in human monocytes exposed to apolar lipid extracts from CDC1551 (ref. 5). This was the first experimental evidence to suggest that strain-related differences in host immune response and long-term disease outcome were directly attributable to unique lipid(s) produced by individual *M. tuberculosis* strains.

We postulated that the hypervirulent phenotype of strain HN878 might be associated with the lipid repertoire of this organism. In an attempt to identify lipids present in HN878 and absent from other

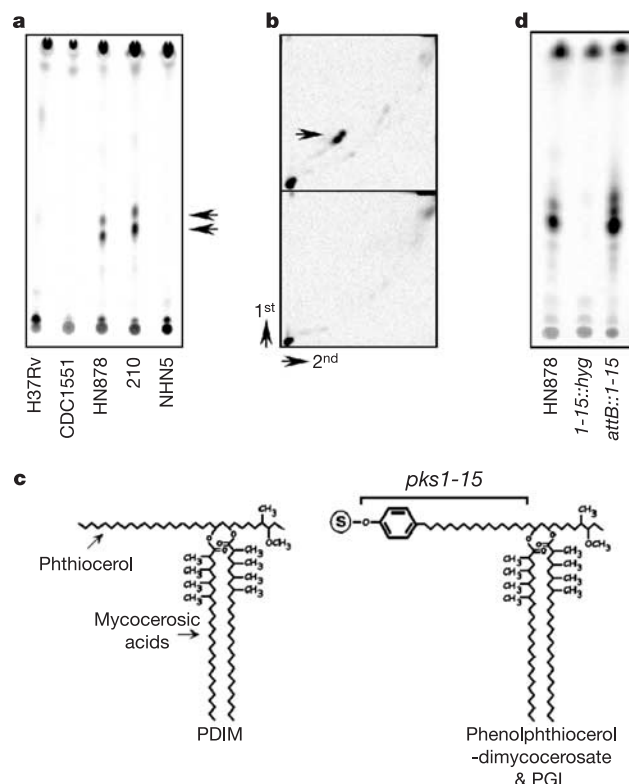


Figure 1 PGL is produced by HN878 and related W-Beijing strains. **a**, Lipids extracted from [¹⁴C]propionic-acid-labelled cultures of the strains indicated were analysed by TLC. The positions of the lipids specific to HN878 and 210 are highlighted (arrows). **b**, Lipids of [¹⁴C]propionic-acid-labelled HN878 (top panel) and pks1-15::hyg (bottom panel) were analysed by two-dimensional TLC, demonstrating loss of the HN878-specific lipids (arrow) in pks1-15::hyg. **c**, Model structures of PDIM and PGL. The region of PGL predicted to be derived from Pks1-15 (ref. 16) and the position of a variable number of carbohydrate residues (S) attached to the phenyl moiety are shown. **d**, TLC of lipids extracted from [¹⁴C]p-hydroxybenzoic-acid-labelled cultures including the complemented attB::pks1-15 strain.

recent clinical isolates of *M. tuberculosis*, various strains were incubated in the presence of [^{14}C]propionic acid to specifically radiolabel lipids possessing methyl branches. Two closely migrating species were identified by thin-layer chromatography (TLC) of apolar extracts of HN878 and W-Beijing strains 210, W10 and W4 (refs 8, 9), but not CDC1551, NHN5 (ref. 2) or the laboratory strain H37Rv¹⁰ (Fig. 1a and data not shown). This result was particularly intriguing because the 210 and W4 strains have also been shown to be associated with a diminished early pro-inflammatory cytokine response within infected human monocytes *in vitro* and hypervirulent behaviour in rodent models of infection¹¹ (G.K., unpublished data). Despite being isolated from distinct geographical regions of the United States, all of the W-Beijing strains examined (including HN878) have virtually indistinguishable IS6110 genotyping patterns¹¹.

The ability to label these lipids with propionate suggested that a polyketide synthase (PKS)-like enzyme capable of extension with methylmalonyl co-enzyme A was involved in their biosynthesis. Several polyketides have recently been described for *M. tuberculosis*, and in many cases represent methyl-branched lipids important for virulence^{12,13}. As part of an ongoing programme to identify

PKS-derived products in *M. tuberculosis*, a set of *pkc* hygromycin-replacement mutants was constructed in HN878, each of which was screened for the absence of the unique lipids identified above. In this way, we established that the HN878 strain disrupted within the

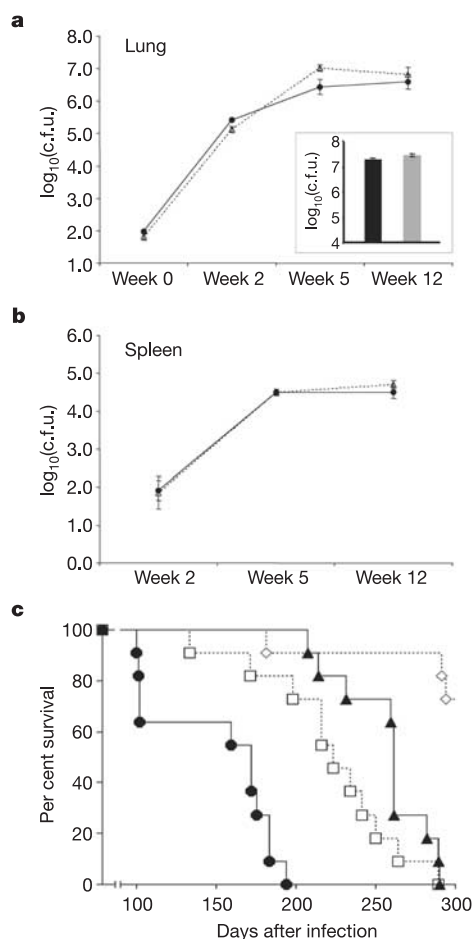


Figure 2 PGL is responsible for the hypervirulent phenotype of HN878 in mice. **a**, Six-week-old B6D2 F₁ mice were infected via aerosol with HN878 (solid line) and *pks1-15::hyg* (dotted line). At 3 h and 2, 5 and 12 weeks after infection lungs were homogenized and plated for c.f.u. determination (3–4 mice per group). Values are $\log_{10}(\text{c.f.u.})$ ($\pm$ s.e.m.). At necropsy, lung c.f.u. were obtained for HN878 (black column) and *pks1-15::hyg* (grey column) (inset). **b**, At 2, 5 and 12 weeks after infection spleens were homogenized and plated. **c**, Survival analysis was carried out for additional mice (11 mice per group) infected with HN878 (filled circles), H37Rv (open squares), *pks1-15::hyg* (filled triangles) and CDC1551 (open diamonds).

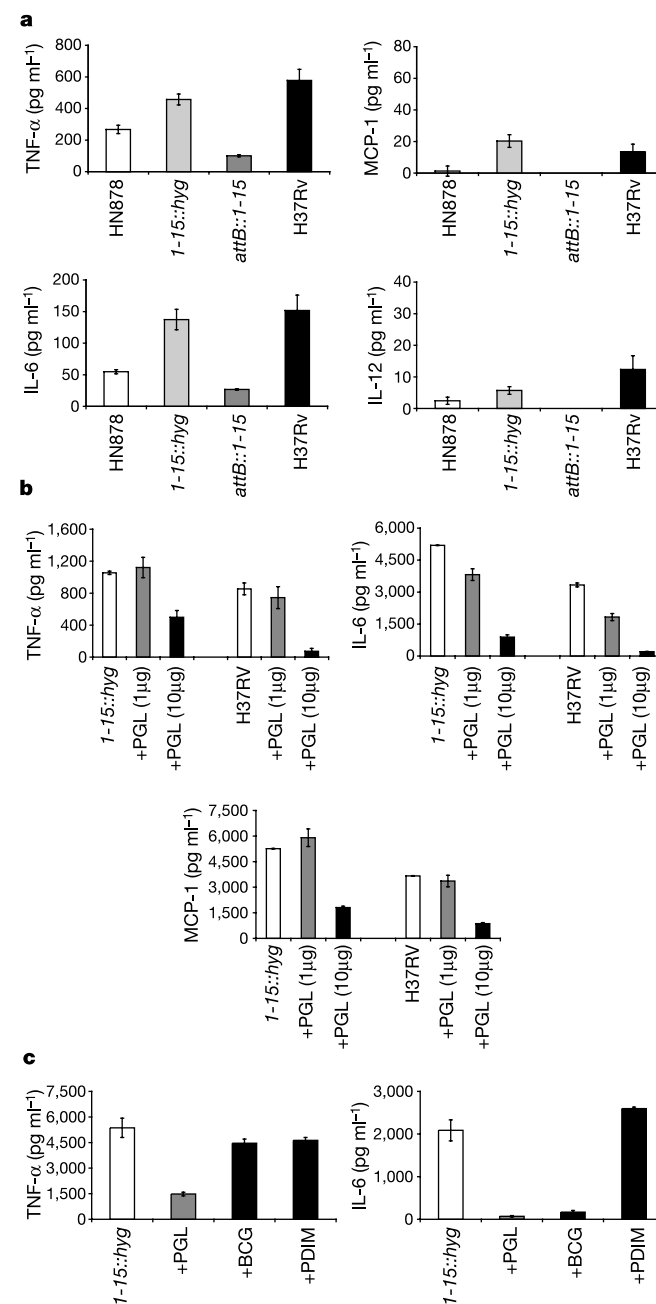


Figure 3 PGL-mediated inhibition of pro-inflammatory cytokine release by murine BMMs. **a**, Culture supernatants obtained 24 h after BMM infections were analysed by ELISA for the presence of the cytokines TNF- α , IL-6, IL-12 and the chemokine MCP-1. Data obtained from one representative experiment performed in triplicate are presented ($\pm$ s.e.m.). **b**, BMMs were incubated for 24 h in the presence of increasing amounts of purified PGL (0, 1 or 10 μ g) in addition to apolar lipids (5 μ g) extracted from HN878 *pks1-15::hyg* and H37Rv. Culture supernatants were analysed by ELISA for TNF- α , IL-6 and MCP-1. IL-12 was unable to be detected in these lipid assays. PGL alone showed no effect in these assays. Representative data ($\pm$ s.e.m.) from a single experiment performed in triplicate are presented. **c**, BMMs were incubated for 24 h in the presence of 2 μ g purified PGL, *M. bovis* (BCG) PGL or PDIM in addition to apolar lipids (1 μ g) extracted from HN878 *pks1-15::hyg*. Culture supernatants were analysed in triplicate ($\pm$ s.e.m.) by ELISA for TNF- α and IL-6.

pksl-15 gene cluster (as annotated in the H37Rv genome¹⁰) no longer produced these lipid species (*pksl-15::hyg*; Fig. 1b). The identity of these lipids was suggested by the genomic context of *pksl-15* and became clear with the publication of the association between *pksl-15* and the production of a complex phenolic glycolipid¹⁴; however, no phenotype had been ascribed to mutants in this pathway.

In H37Rv and CDC1551, the absence of PGL production was confirmed to be associated with the previously reported 7-base-pair deletion resulting in a frameshift in what would otherwise be predicted to be a single gene encompassing *pksl* and *pksl5* (refs 14, 15). An intact *pksl-15* gene was identified for HN878, 210, W4 and W10 (data not shown) and, as predicted, each of these strains produces the unique lipids. Interestingly, the unrelated NHN5 strain fails to produce these lipids (Fig. 1a), yet doesn't contain the lesion resulting in their absence from H37Rv and CDC1551. This suggests that multiple independent genetic alterations may have led to the loss of production of such lipids, and may imply that their production is selected against within the existing population of *M. tuberculosis* strains.

pksl-15 is located in a region of the genome that includes the *mas*, *mmpL7* and *ppsA-E* genes, all of which are known to be involved in the formation or transport of phthiocerol dimycocerosate (PDIM). PDIM is a complex wax located in the outer wall of mycobacteria, loss of which has been associated with attenuation of *M. tuberculosis*^{12,16}. Importantly, PDIM synthesis and export were found to be unaffected in HN878 *pksl-15::hyg* (data not shown). A class of lipids structurally related to PDIM and collectively referred to as phenolic glycolipids (Fig. 1c) have been reported to be present in *Mycobacterium bovis*, *Mycobacterium leprae* and other pathogenic mycobacteria, including a limited number of *M. tuberculosis* isolates^{14,17}. Labelling of HN878 with [¹⁴C]*p*-hydroxybenzoic acid, a known precursor of PGL¹⁶, resulted in radioactive species co-migrating with the HN878-specific lipids labelled with [¹⁴C]propionate, confirming that the lipids contain a *p*-hydroxybenzyl constituent (Fig. 1d). ¹H-NMR spectroscopy (data not shown) was carried out on material purified by preparative TLC, confirming the presence of multiple sugars (chemical shift (δ) = 5.13 and 5.50 parts per million (p.p.m.)), phenolic protons (δ = 6.9–7.1 p.p.m.) and an esterified β -diol (δ = 4.84 p.p.m.), all consistent with the known structure of PGL¹⁸. Inactivation of *MmpL7*, a transmembrane protein shown to facilitate PDIM transport¹², blocked not only PDIM secretion as previously reported but also PGL secretion. Although the ratio of the individual PGLs is slightly altered, PGL synthesis is otherwise unaffected in this *MmpL7*-inactivation mutant, suggesting that attachment of sugar residues occurs before transport across the cell membrane (data not shown).

To address the role of PGL in tuberculosis infection and its association with the hypervirulent phenotype displayed by HN878 (ref. 6), we performed low-dose aerosol infections of BL6/DBA2 F₁ mice (50–100 colony-forming units (c.f.u.) per mouse) with HN878 and HN878 *pksl-15::hyg*. There was no significant difference in bacterial numbers within the lung over the initial 12 weeks of infection, yet there was an almost 90-day difference in median survival between the two groups ($P < 0.0001$) (Fig. 2). Indeed, the survival of mice infected with *pksl-15::hyg* was not significantly different to that of the intermediately virulent H37Rv^{5,6} ($P = 0.07$). No alteration in bacterial number in the spleen was observed, indicating that absence of PGL has no detrimental effect on bacterial dissemination in this model. Enumeration of bacteria at the time of necropsy confirmed that the difference in mouse survival is not attributable to differences in bacterial survival or bacillary load. A second set of infections carried out with a larger infecting inoculum (300–400 c.f.u. per mouse) confirmed the *pksl-15::hyg* phenotype. Despite the higher antigenic load in this case, a difference in median survival of 42 days ($P = 0.004$) was still observed (data not shown).

Diverse biological activities ranging from suppression of human

monocytic responses *in vitro* to a role in Schwann cell invasion have been attributed to the abundant PGL-1 of *M. leprae*^{19–24}. To explore the mechanism by which the production of PGL contributes to the hyperlethality observed in mice infected with HN878, we compared pro-inflammatory cytokine production in bone-marrow-derived macrophages (BMMs) infected with HN878, *pksl-15::hyg* and H37Rv. In addition, we included in these studies an HN878 *pksl-15::hyg* strain containing an intact copy of *pksl-15* under the control of the *hsp60* promoter on the integrative expression plasmid pMV361 (ref. 25). Examination of this strain, *attB::pksl-15*, confirmed that expression of *pksl-15* could fully restore PGL production and function to the HN878 *pksl-15::hyg* strain. On the basis of the incorporation of [¹⁴C]*p*-hydroxybenzoic acid, production of PGL within *attB::pksl-15* was three times greater than that of the parental HN878 strain (Fig. 1d). Infection of BMMs revealed an inverse correlation between PGL production and the level of secretion of the pro-inflammatory mediators tumour-necrosis factor- α (TNF- α , interleukin (IL)-6, IL-12 and MCP-1 (monocyte chemoattractant protein-1) (Fig. 3a). When comparing HN878 with the *pksl-15::hyg* mutant, the degree of inhibition due to the presence of PGL was typically in the order of 2–2.5-fold, depending on the particular cytokine assayed. Moreover, this PGL-mediated inhibition of the innate pro-inflammatory response is reflected in the poor survival of mice infected with wild-type HN878 (Fig. 2c). The cytokines IL-4, IL-10 and IL-13, often linked with TH₂-type allergic responses and an alternative pathway of macrophage activation²⁶, were not produced at detectable levels in these assays. Finally, we incubated BMMs with apolar lipids extracted from either *pksl-15::hyg* or H37Rv in the presence of increasing amounts of purified PGL. These experiments revealed a dose-dependent inhibition of pro-inflammatory cytokine release similar to that observed for BMMs infected with the PGL-producing strains HN878 and *attB::pksl-15* (Fig. 3b). The specificity of the response is highlighted by the fact that the structurally related PGL from *M. bovis* (BCG) bearing only a single carbohydrate substituent¹⁷ showed differential effects in terms of TNF- α and IL-6 secretion, whereas PDIM displayed no inhibitory activity (Fig. 3c).

PGL confers on a subset of *M. tuberculosis* isolates the ability to inhibit the release of key inflammatory effector molecules by cells of the host's innate immune response. Ultimately, this inhibition manifests itself in the hyperlethal phenotype observed in mice. So far, the link between PGL and hypervirulence is restricted to an experimental model that in many ways does not adequately reflect the pathology observed in human tuberculosis¹. Consequently, the role of PGL in human granulomatous tuberculosis disease remains to be determined. Of particular importance will be a careful examination of any linkage between PGL and the infection sequelae associated with the W-Beijing family, notably multidrug resistance, dissemination and outbreaks of disease^{8,9}. However, the results herein provide a clear demonstration that the spectrum of tuberculosis disease is likely to reflect not only variable host factors, but also the variable expression of bacterial factors including PGL. □

Methods

Bacterial strains and media

M. tuberculosis CDC1551 (T. Shinnik), H37Rv (American Type Culture Collection 27294), HN878 (J. Musser), NHN5 (J. Musser), W4 (TB Center collection) and 210 (D. Cave) were cultured using 7H9 liquid medium (Difco) containing 10% ADC (albumin dextrose complex), 0.5% glycerol and 0.1% Tween-80 or 7H11 agar containing 10% OADC (oleic acid plus ADC), 0.5% glycerol and 0.1% L-asparagine. Where necessary, media were supplemented with 50 $\mu\text{g ml}^{-1}$ hygromycin or 25 $\mu\text{g ml}^{-1}$ kanamycin.

Mutant construction

Using sense (5'-CAGATCTCGGCTGCTCAAT-3') and antisense (5'-GAGCCAGGACTGCAACACCT-3') primers specific for the H37Rv *pksl* sequence, a 1.9-kb fragment was amplified from H37Rv genomic DNA and cloned into pCR-Blunt (Invitrogen). The resultant plasmid was restricted with *Bst*EI and blunt-end ligated with the 1.3-kb hygromycin expression cassette from p16R1 (ref. 27). The *pksl::hyg* fragment was released from pCR-Blunt by digestion with *Eco*Rv and *Spe*I and inserted via blunt-end ligation into

BamHI-restricted pPR23-1 (ref. 28). Similarly, a 2.3-kb fragment was amplified using sense (5'-ATAAGATCGCGCCGCGTGTGTCGAGCAATATCGACAG-3') and antisense (5'-GGACTAGTCCAGAGCCACTACCGTGAGCAAG-3') primers specific for the H37Rv *mmpL7* gene, digested with *NorI*, *SpeI* and cloned into pcDNA2.1 (Invitrogen). A 1.3-kb *NorI* fragment containing the hygromycin cassette of pHint²⁹ was then inserted into the *SmaI* site of the *mmpL7* fragment. The 3.6-kb *NorI/SpeI* fragment containing *mmpL7::hyg* was excised from pcDNA2.1 and inserted into the same sites of pPR23-1 (ref. 28). Transformation and selection procedures were as described previously²⁸. To generate *attB::pks1-15*, we initially sub-cloned a 6.5-kb *XbaI/XmnI* fragment containing *pks1-15* from the appropriate H37RV BAC clone (provided by R. Brosch) into *EcoRV/XbaI*-restricted pBluescript II SK (Stratagene). The resultant plasmid was digested with *NorI/BamHI* into which the HN878 *pks15* sequence (including the *pks1-15* junction) generated by polymerase chain reaction (PCR) was inserted. The entire *pks1-15* insert was excised with *DraI/HindIII*, cloned into pMV361 (ref. 25) and transformed into HN878 *pks1-15::hyg*.

Biochemical analysis

Cultures were grown to an absorbance of 0.25 at 650 nm (A_{650}) at which point $0.1 \mu\text{Ci ml}^{-1}$ [$1\text{-}^{14}\text{C}$]propionic acid or $0.7 \mu\text{Ci ml}^{-1}$ [ring- ^{14}C (U)]4-hydroxybenzoic acid (American Radiolabelled Chemicals) were added and incubated for a further 48 and 96 h, respectively. Cell pellets were re-suspended in MeOH:0.3% NaCl (aq.) (10:1) and apolar lipids extracted twice with petroleum ether as described³⁰. Filtered (0.2 μM) culture supernatants were extracted twice with 0.5 volumes of petroleum ether. Lipid extracts were analysed on Silica Gel-60 TLC plates (Merck) and developed in $\text{CHCl}_3\text{:MeOH}$ (94:6). For two-dimensional TLC, plates were developed in $\text{CHCl}_3\text{:MeOH}$ (95:5) (1st dimension) and toluene:acetone (70:30) (2nd dimension). TLC plates were visualized using a Storm 860 PhosphorImager (Molecular Dynamics). For *in vitro* studies, approximately 1 mg of PGL was extracted and purified from 2-l cultures of HN878 by preparative Silica Gel TLC. One-dimensional ^1H -NMR spectroscopy was carried out on the purified material using a Mercury 300 MHz instrument and VNMR 6.1c software (Varian) with CDCl_3 as solvent and tetramethylsilane as an internal reference.

Animal studies

Before infection, *M. tuberculosis* cultures were adjusted to A_{650} of 0.5 and stored at -70°C as 20% glycerol stocks. Inocula were prepared by diluting these stocks to 4×10^6 c.f.u. ml^{-1} in PBS/Tween-80 (0.05%). Six-week-old B6D2 F₁ (C57BL/6 $\times$ DBA/2; F₁ progeny) mice (Taconic) were infected (30 mice per group) via aerosol for 10 min using a BioAerosol Nebulizing Generator (CH Technologies). In this manner, approximately 50–100 c.f.u. per lung were implanted as confirmed by homogenizing lungs (4 mice per group) in 7H9/ADC at 3 h after infection and plating for c.f.u. determination. Additional mice were killed at 2, 5 and 12 weeks after infection and serial dilutions of lung and spleen homogenates were plated. The log-rank test was used to determine statistical significance of survival differences observed in mice (GraphPad Prism v3.0).

Cytokine analysis

To generate BMMs, bone marrow flushed from mouse femurs (B6D2 F₁) was cultured for 7 days in high glucose DMEM (Gibco) supplemented with 20% FCS, 1 mM pyruvate, 2 mM glutamine and 30% 1929 conditioned medium. 10^6 BMMs (in high glucose DMEM plus 10% FCS, 1 mM pyruvate and 2 mM glutamine) seeded into 24-well plates were infected with various *M. tuberculosis* strains at a ratio of 1:1 for 4 h after which extracellular bacteria were removed by repeated washing with PBS. Culture supernatants were removed 18–24 h after infection, centrifuged and analysed for the presence of TNF- α , IL-6, IL-12 and MCP-1 by enzyme-linked immunosorbent assay (ELISA) (R&D Systems). Infected cells were lysed with 0.025% SDS and serial dilutions plated for c.f.u. determinations. Purified PGL (1 μg or 10 μg) and crude apolar lipid extracts (5 μg) prepared from HN878 *pks1-15::hyg* or H37Rv were re-suspended in petroleum ether, added to 24-well tissue culture plates and the solvent allowed to evaporate before adding BMMs as above. Additionally, similar assays were carried out in the presence of purified PDIM or *M. bovis* (BCG) PGL. Each infection or lipid treatment was performed in triplicate with macrophages prepared from at least three individual mice.

Received 7 May; accepted 13 July 2004; doi:10.1038/nature02837.

1. Tiruvilumala, P. & Reichman, L. B. Tuberculosis. *Annu. Rev. Publ. Health* **23**, 403–426 (2002).
2. Sreevatsan, S. *et al.* Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl Acad. Sci. USA* **94**, 9869–9874 (1997).
3. Fleischmann, R. D. *et al.* Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriol.* **184**, 5479–5490 (2002).
4. North, R. J., Ryan, L., LaCourse, R., Mogues, T. & Goodrich, M. E. Growth rate of mycobacteria in mice as an unreliable indicator of mycobacterial virulence. *Infect. Immun.* **67**, 5483–5485 (1999).
5. Manca, C. *et al.* *Mycobacterium tuberculosis* CDC1551 induces a more vigorous host response *in vivo* and *in vitro*, but is not more virulent than other clinical isolates. *J. Immunol.* **162**, 6740–6746 (1999).
6. Manca, C. *et al.* Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN- α/β . *Proc. Natl Acad. Sci. USA* **98**, 5752–5757 (2001).
7. Valway, S. E. *et al.* An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. *N. Engl. J. Med.* **338**, 633–639 (1998).
8. Bifani, P. J., Mathema, B., Kurepina, N. E. & Kreiswirth, B. N. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol.* **10**, 45–52 (2002).
9. Glynn, J. R., Whiteley, J., Bifani, P. J., Kremer, K. & van Soolingen, D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerg. Infect. Dis.* **8**, 843–849 (2002).

10. Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
11. Manca, C. *et al.* Differential monocyte activation underlies strain specific *M. tuberculosis* pathogenesis. *Infect. Immun.* (in the press).
12. Cox, J. S., Chen, B., McNeil, M. & Jacobs, W. R. Jr Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature* **402**, 79–83 (1999).
13. Sirakova, T. D., Thirumala, A. K., Dubey, V. S., Sprecher, H. & Kolattukudy, P. E. The *Mycobacterium tuberculosis* *pks2* gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J. Biol. Chem.* **276**, 16833–16839 (2001).
14. Constant, P. *et al.* Role of the *pks15/1* gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the *pks15/1* gene. *J. Biol. Chem.* **277**, 38148–38158 (2002).
15. Marmiesse, M. *et al.* Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiol.* **150**, 483–496 (2004).
16. Kolattukudy, P. E., Fernandes, N. D., Azad, A. K., Fitzmaurice, A. M. & Sirakova, T. D. Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol. Microbiol.* **24**, 263–270 (1997).
17. Vergne, I. I. & Daffe, M. Interaction of mycobacterial glycolipids with host cells. *Front. Biosci.* **3**, 865–876 (1998).
18. Hunter, S. W. & Brennan, P. J. A novel phenolic glycolipid from *Mycobacterium leprae* possibly involved in immunogenicity and pathogenicity. *J. Bacteriol.* **147**, 728–735 (1981).
19. Mehra, V., Brennan, P. J., Rada, E., Convit, J. & Bloom, B. R. Lymphocyte suppression in leprosy induced by unique *M. leprae* glycolipid. *Nature* **308**, 194–196 (1984).
20. Fournie, J.-J., Adams, E., Mullins, R. J. & Basten, A. Inhibition of human lymphoproliferative responses by mycobacterial phenolic glycolipids. *Infect. Immun.* **57**, 3653–3659 (1989).
21. Vachula, K., Holzer, T. J. & Andersen, B. R. suppression of monocyte oxidative responses by phenolic glycolipid 1 of *Mycobacterium leprae*. *J. Immunol.* **142**, 1696–1701 (1989).
22. Silva, C. L., Faccioli, L. H. & Foss, N. T. Suppression of human monocyte cytokine release by phenolic glycolipid-1 of *Mycobacterium leprae*. *Int. J. Lepr.* **61**, 107–108 (1993).
23. Hashimoto, K. *et al.* *Mycobacterium leprae* infection in monocyte-derived dendritic cells and its influence on antigen-presenting function. *Infect. Immun.* **70**, 5167–5176 (2002).
24. Ng, V. *et al.* Role of the cell wall phenolic glycolipid-1 in the peripheral nerve predilection of *Mycobacterium leprae*. *Cell* **103**, 511–524 (2000).
25. Stover, C. K. *et al.* New use of BCG for recombinant vaccines. *Nature* **351**, 456–460 (1991).
26. Gordon, S. Alternative activation of macrophages. *Nature Rev. Immunol.* **3**, 23–35 (2003).
27. Garbe, T. R. *et al.* Transformation of mycobacterial species using hygromycin resistance as selectable marker. *Microbiol.* **140**, 133–138 (1994).
28. Pelicic, V. *et al.* Efficient allelic exchange and transposon mutagenesis in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **94**, 10955–10960 (1997).
29. O'Gaora, P. *et al.* Mycobacteria as immunogens: development of expression vectors for use in multiple mycobacterial species. *Med. Princ. Prac.* **6**, 91–96 (1997).
30. Slayden, R. A. & Barry, C. E. III in *Mycobacterium tuberculosis protocols* (eds Parish, T. & Stoker, N. G.) 229–245 (Humana, New Jersey, 2001).

Acknowledgements The authors wish to thank J. Gonzales and M. Goodwin for their assistance with animal studies and NMR spectroscopy, respectively. G.K. is supported by grants from the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.E.B. (clifton_barry@nih.gov).

Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*

Tzvi Tzfira, Manjusha Vaidya & Vitaly Citovsky

Department of Biochemistry and Cell Biology, State University of New York, Stony Brook, New York 11794-5215, USA

Genetic transformation of plant cells by *Agrobacterium* represents a unique case of trans-kingdom DNA transfer¹. During this process, *Agrobacterium* exports its transferred (T) DNA and several virulence (Vir) proteins into the host cell², within which T-DNA nuclear import is mediated by VirD2 (ref. 3) and VirE2 (ref. 4) and their host cell interactors ATKAP- α ⁵ and VIP1 (ref. 6), whereas its integration is mediated mainly by host cell proteins^{7–9}.

The factors involved in the uncoating of T-DNA from its cognate proteins, which occurs before integration into the host genome, are still unknown. Here, we report that VirF—one of the few known exported Vir proteins whose function in the host cell remains unknown—is involved in targeted proteolysis of VIP1 and VirE2. We show that VirF localizes to the plant cell nucleus and interacts with VIP1, a nuclear protein. VirF, which contains an F-box motif¹⁰, significantly destabilizes both VIP1 and VirE2 in yeast cells. Destabilization of VIP1 in the presence of VirF was then confirmed *in planta*. These results suggest that VIP1 and its cognate VirE2 are specifically targeted by the VirF-containing Skp1–Cdc53–cullin–F-box complex for proteolysis. The critical role of proteasomal degradation in *Agrobacterium*-mediated

genetic transformation was also evident from inhibition of T-DNA expression by a proteasomal inhibitor. *Agrobacterium* genetically transforms eukaryotic cells from diverse origins, from its natural plant hosts (reviewed in refs 8, 9) to yeast¹¹ to human¹². During this process, the bacterial T-DNA is thought to be imported into the host cell nucleus as a DNA–protein complex (T-complex) composed of a single T-DNA strand packaged by numerous VirE2 molecules, a single VirD2 molecule attached to its 5' end, and VIP1 proteins bound to VirE2 (reviewed in refs 13, 14). Before integration into the host genome, the T-DNA must be uncoated from its cognate proteins by a process yet to be determined. We suggest that this is achieved by means of targeted proteolysis mediated by the bacterial virulence protein VirF, which

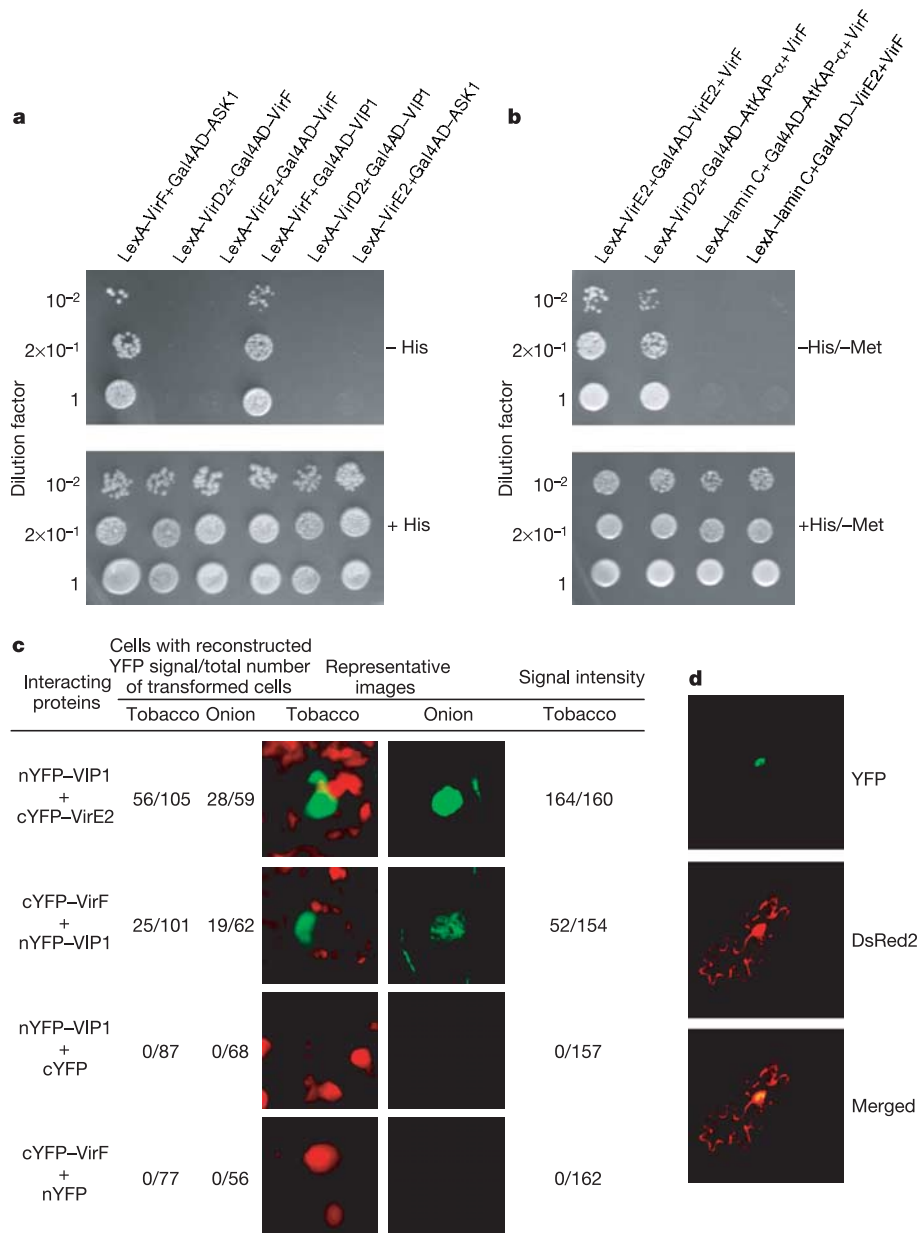


Figure 1 Specific interaction between VirF and VIP1 in yeast and plant cells. **a, b**, Yeast two-hybrid assay. L40(–ura3) cells were grown in histidine-deficient medium with 10 mM of 3-amino-1,2,4-triazole (–His) or in histidine-containing medium (+His); free VirF was co-expressed in the absence of methionine (–Met) from a methionine-repressible promoter. **c**, *In planta* BiFC assay. Signal intensity of reconstructed YFP

(green) is presented relative to plastid autofluorescence (red). **d**, Nuclear localization of the interacting VirF and VIP1 proteins was confirmed in tobacco cells by co-localization (yellow colour in the merged image) of reconstructed YFP (yellow) with free DsRed2 (red), which labels both the cytoplasm and the nucleus²². All images are single confocal sections.

is exported into plant cells² and contains an F-box motif that binds the plant homologue of Skp1, ASK1 (ref. 10). Since its discovery as a host range factor^{15–17}, the molecular basis for the function of VirF has remained elusive. We reasoned that VirF might be involved in destabilization of the protein components of the T-complex by participating in a Skp1–Cdc53–cullin–F-box (SCF) complex involved in protein degradation (reviewed in ref. 18). In this model, one or more of the T-DNA-associated proteins would be recognized by VirF and represent a substrate for targeted proteolysis. Figure 1a shows that, in the yeast two-hybrid system, VirF, which also bound ASK1, specifically interacted with VIP1, but not with VirD2 or VirE2. This latter lack of interaction was not due to instability of VirE2 and VirD2 in the presence of VirF because co-expression of free VirF did not affect VirE2–VirE2 (ref. 19) and VirD2–AtKAP- α ⁵ interactions (Fig. 1b). VIP1, known to bind VirE2 (ref. 6), did not interact with VirD2, and no interaction between ASK1 and VirE2 was detected. Thus, of all the known T-complex protein components, VirF specifically recognizes VIP1 in yeast.

The VirF–VIP1 interaction was confirmed *in planta* using the bimolecular fluorescence complementation (BiFC) assay²⁰. In this approach, the yellow fluorescent protein (YFP) molecule is split into amino-terminal (nYFP) and carboxy-terminal (cYFP) non-fluorescent portions. Restoration of the YFP fluorescence signal is achieved when nYFP and cYFP are brought together as fusions with interacting proteins²⁰. Figure 1c shows that YFP fluorescence was restored in tobacco and onion cells—both of which are infected by *Agrobacterium*²¹—after co-expression of nYFP–VIP1 and cYFP–VirF, yet the intensity of the reconstructed YFP signal was weaker

and the number of cells exhibiting the fluorescence signal was generally lower than those observed for the VIP1–VirE2 interaction. These differences probably reflect the biological effect of the VirF–VIP1 interaction (see below). No YFP fluorescence was detected when nYFP–VIP1 or cYFP–VirF were co-expressed with unfused cYFP or nYFP, respectively (Fig. 1c); this lack of background signal, previously reported for BiFC in animal cells²⁰, facilitated identification of cells exhibiting the reconstructed YFP fluorescence.

If VirF and ASK1 are involved in uncoating of the T-DNA before its expression and integration, they both should reside in the host cell nucleus where VirF would interact with its target, VIP1. Indeed, green fluorescent protein (GFP)-tagged VirF and ASK1 accumulated in the nuclei of plant cells (data not shown). Furthermore, the VirF–VIP1 complexes were also intranuclear, co-localizing with co-expressed DsRed2 (Fig. 1d), which labels both the nucleus and the cytoplasm²².

Next, the VirF-mediated protein destabilization was assayed by a previously reported approach in which SCF-dependent protein degradation is monitored using fusions to a reporter protein; decrease in the reporter activity indicates decreased stability²³. We examined the effect of VirF on the stability of three protein components of the T-complex (VIP1, VirE2 and VirD2) in two experimental systems: yeast and plant. In yeast, expression of GFP-tagged VIP1, VirE2 and VirD2 was controlled by a galactose-inducible promoter, whereas expression of unlabelled VirF was controlled by a methionine-repressible promoter. We compared the GFP signal (black bars in Fig. 2) produced in control systems (that is, after cell growth in galactose-containing medium) with the GFP signal (grey bars in Fig. 2) accumulated after transferring the cells to a glucose-containing medium, with or without induction of VirF expression. Figure 2a shows that VirF significantly depleted the amount of GFP–VIP1, reducing it to about 25% of its original level. No such changes in GFP signal were observed under VirF-repressing conditions (Fig. 2a), indicating the specific role of VirF in destabilization of GFP–VIP1. Furthermore, VirF expression also caused destabilization of up to 60% of GFP–VirE2 when co-expressed with free, unfused VIP1 (Fig. 2b). This destabilization was VIP1-dependent because GFP–VirE2 co-expressed with VirF in the absence of VIP1 remained stable (Fig. 2b). GFP–VirD2, which does not interact with VirF (see Fig. 1a) or VIP1 (ref. 6), was not destabilized when co-expressed with VirF and VIP1 (Fig. 2c). Note that, under our experimental conditions, the VirF-induced destabilization of VIP1 and VirE2 was incomplete, explaining why interactions between VirF and VIP1 could still be detected in the two-hybrid system and in the BiFC assay; most likely, the residual levels of these proteins were sufficient to induce expression of the two-hybrid reporter genes and to reconstruct detectable levels of the YFP signal (see Fig. 1a, c). That these low amounts of VirE2 and VIP1 still remained in the cells may be due to their *de novo* synthesis or recalcitrance of some of the expressed protein to degradation.

Because VirF is an F-box protein thought to participate in the SCF-mediated degradation¹⁰, we directly tested whether the VirF-containing SCF (SCF^{VirF}) pathway is involved in destabilization of GFP–VIP1. To this end, we used a yeast temperature-sensitive mutant in the Skp1 component of the SCF complex, *skp1-4* (ref. 24). Figure 2d shows that VirF expressed in *skp1-4* cells substantially destabilized GFP–VIP1 at the permissive temperature (25°C), but not after shift to the restrictive temperature (37°C). As expected, in the parental, wild-type yeast strain, destabilization of GFP–VIP1 occurred with equal efficiency at both temperatures (not shown).

Collectively, these results suggest that VirF specifically destabilizes VIP1 and VirE2 via targeted proteolysis mediated by SCF^{VirF}. Also, because VirF binds VIP1 but not VirE2 (see Fig. 1a), its effect on VIP1 stability was probably direct whereas the stability of VirE2 was affected indirectly, presumably through the VIP1–VirE2 interaction.

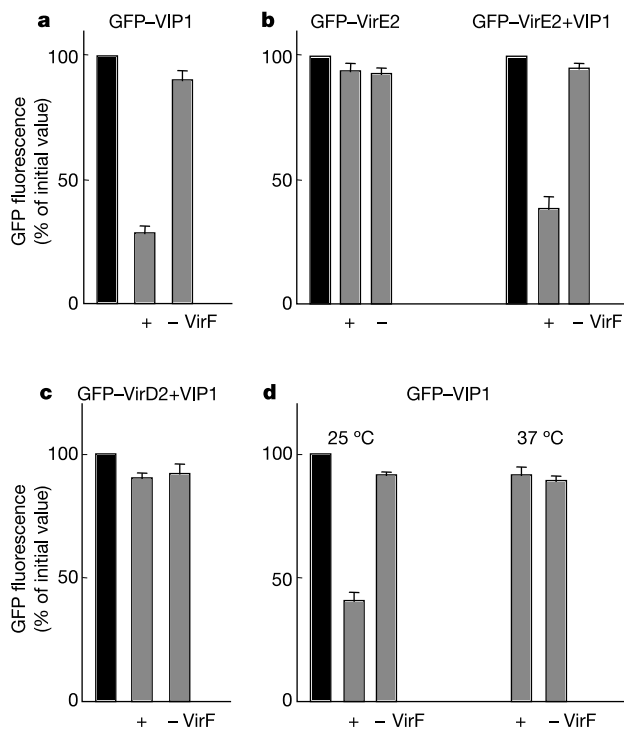


Figure 2 VirF-mediated and Skp1-dependent destabilization of VIP1 and VirE2 in yeast cells. **a**, Destabilization of GFP–VIP1. **b**, Destabilization of GFP–VirE2. **c**, Lack of destabilization of GFP–VirD2. **d**, Destabilization of GFP–VIP1 at permissive (25°C) but not at restrictive (37°C) temperature in *skp1-4* cells. Combinations of target genes are indicated on the top of each panel, and initial levels of GFP fluorescence, measured after induction of the target proteins in the absence of VirF expression, are indicated by black bars. Fluorescent signal measured after induction (+) or continued repression (–) of VirF is indicated by grey bars. Bars represent mean $\pm$ standard error ($n = 5$).

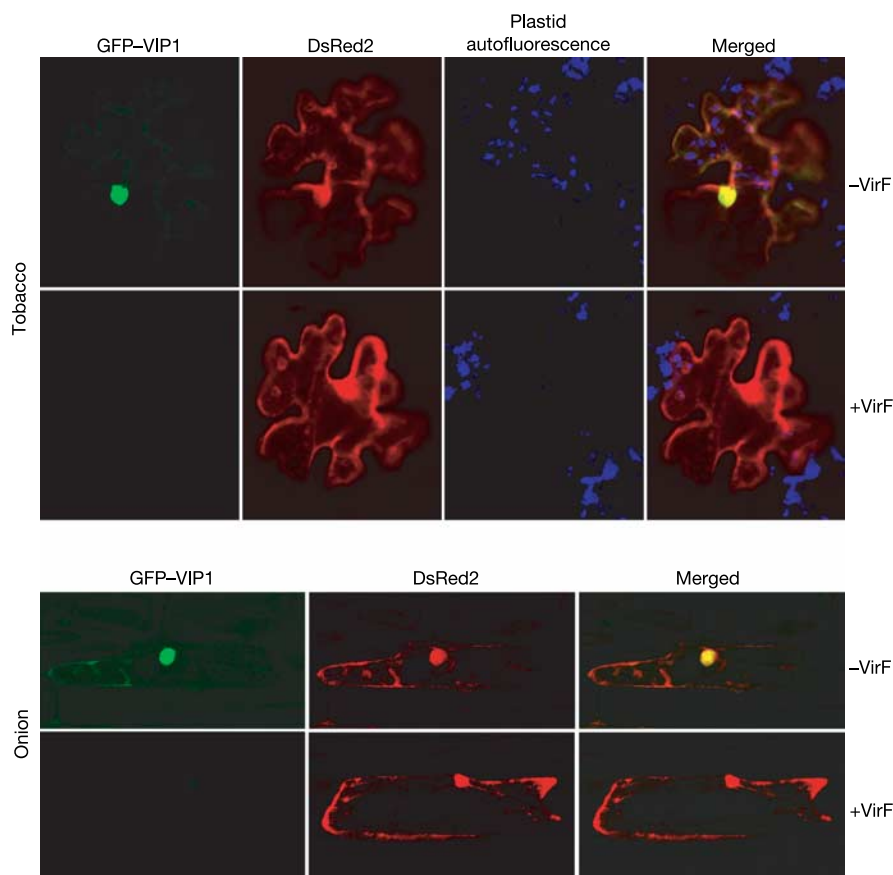


Figure 3 VirF-mediated destabilization of VIP1 *in planta*. Plant tissues were transiently transformed with pGDR-GFP-VIP1 alone or mixed with pRTL2-VirF. pGDR-GFP-VIP1 expresses both GFP-VIP1 and free DsRed2 whereas pRTL2-VirF expresses unfused VirF.

DsRed2 labels the cytoplasm and the nucleus and identifies transformed cells. GFP, green; DsRed2, red; co-localizing merged GFP and DsRed2, yellow; plastid autofluorescence, blue. All images are single confocal sections.

The effect of VirF on VIP1 stability was then examined directly *in planta* by transiently co-expressing GFP-VIP1 and free VirF. The DsRed2 fluorescent marker, located on the same plasmid that expressed GFP-VIP1, was used to identify the transformed cells and to visualize their nucleus and cytoplasm, between which it is known to partition²². As expected^{6,25}, in the absence of VirF, VIP1 predominantly accumulated in the nuclei of tobacco and onion cells (Fig. 3). In the presence of VirF, however, a significant population of the transformed cells displayed no GFP-VIP1 signal, but still accumulated the DsRed2 fluorescence (Fig. 3). Thus, although the DsRed2/GFP-VIP1-encoding construct had been delivered to and was expressed in these cells, GFP-VIP1 was depleted to levels below our detection limits. In addition, consistent with the reconstructed but weak fluorescence signal in the BiFC experiments (see Fig. 1c, d), low, residual levels of GFP-VIP1 fluorescence could still be observed in some cells co-expressing GFP-VIP1 and VirF (not shown). These data suggest that, in plant cells, the presence of VirF destabilizes VIP1, potentially through the SCF^{VirF} pathway. Owing to technical difficulties with reproducible expression of three co-bombarded constructs, we were unable to examine the effect of VirF on the stability of VirE2 in the presence of overexpressed VIP1.

To examine the effect of the 26S proteasome on T-DNA expression, tobacco tissues were treated with the proteasome inhibitor MG132 (ref. 26) for 2 h before and during the first 2 h of their inoculation with *Agrobacterium*, and T-DNA expression was monitored by polymerase chain reaction with reverse transcription (RT-PCR)²⁷. Figure 4a shows that, in untreated leaf disks, T-DNA-specific transcripts were detected already 2 h after inoculation, and that addition of DMSO, the MG132 solvent, to the incubation

medium did not alter the efficiency and the pattern of T-DNA expression.

In contrast, RT-PCR analysis of the MG132-treated disks detected the T-DNA-specific products significantly later, 6–8 h after inoculation (Fig. 4a). This inhibitory effect of MG132 on T-DNA expression was observed with both concentrations (25 μ M and 50 μ M) of the proteasome inhibitor used in our experiments, with the higher amount of MG132 being slightly more effective (Fig. 4a). Thus, MG132 affected T-DNA expression early in the infection process. This inhibition of T-DNA expression was specific, as MG132 had no discernible effect on expression of genes introduced into plant tissues by an *Agrobacterium*-independent method. Indeed, Fig. 4b shows that no differences in expression of the *gfp* gene were detected when this gene was bombarded into untreated leaf disks, or into disks treated with DMSO or with 25 μ M or 50 μ M of MG132, indicating that MG132 did not simply interfere with the gene expression ability of the plant cell. Although the effect of MG132 on T-DNA expression may still be also due to other, more indirect effects on the plant cell, these results, taken together with VirF-induced destabilization of VirE2 and VIP1, strongly suggest involvement of proteasomal degradation in *Agrobacterium*-mediated genetic transformation.

VirE2 is thought to package the transported T-DNA molecule into a T-complex, protect it and, together with the host factor VIP1, assist its nuclear import (reviewed in refs 13, 14). Thus, disassembly of the T-complex of VirE2 and VIP1 probably occurs within the nucleus, before integration. Our data provide direct evidence for the role of VirF in intranuclear proteolysis of two of the protein components of the *Agrobacterium* T-complex—VIP1 and VirE2—

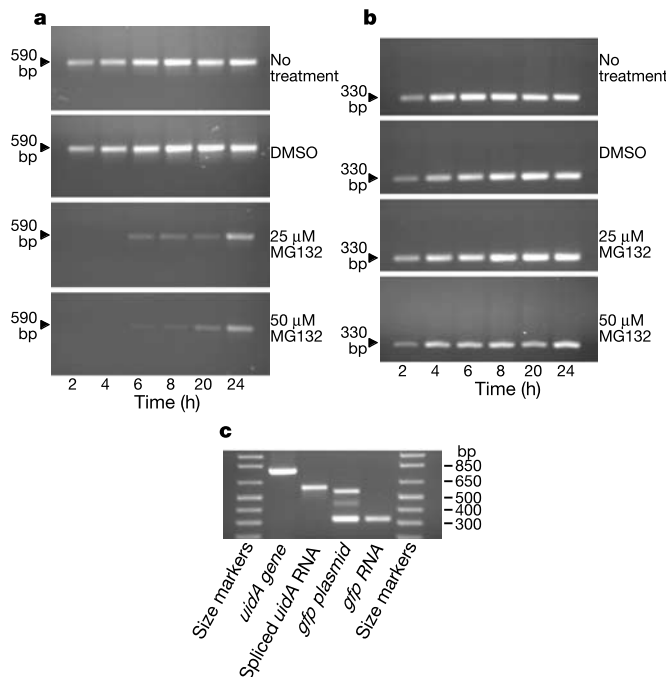


Figure 4 The effect of 26S proteasome inhibition on T-DNA expression. **a, b**, RT-PCR analysis of T-DNA (**a**) and bombarded *gfp* (**b**) expression in MG132-treated leaf disks. **c**, Control PCR reactions. T-DNA expression was detected using *uidA*-specific primers that yield a 590-base-pair (bp) PCR product for the spliced *uidA* transcript and a 779-bp PCR product for the intron-containing *uidA* sequence. *gfp* expression was detected using *gfp*-specific primers, in combination with a third primer located within the promoter of the *gfp* gene, that yield a 330-bp PCR product for the *gfp* transcript and a 560-bp PCR product for the *gfp* promoter-containing plasmid sequences.

via the SCF^{VirF}-mediated pathway. This notion is supported by our observations that early expression of T-DNA, which requires uncoating of the T-complex, is largely blocked by an inhibitor of the 26S proteasome. Interestingly, as a host range factor, VirF is required for transformation of some but not all plant species^{15–17,28,29}. Thus, in plants that do not require VirF, its function may be fulfilled by an as yet unidentified nuclear-localized cellular F-box protein; indeed, many other pathogenic bacteria export to their eukaryotic hosts proteins that mimic functions of the host cell factors required for the infection (reviewed in ref. 30). □

Methods

BiFC assay for protein–protein interaction in planta

VIP1, VirE2 and VirF were fused to the C terminus of either the 5' half (nYFP) or the 3' half (cYFP) of the YFP open reading frame dissected between codons 154 and 155 (ref. 20). Different combinations of the tested proteins were mixed and co-bombarded into tobacco leaves or onion scales. Transformation efficiency was estimated by parallel bombardment of full-length YFP. Signal intensity of reconstructed YFP is presented relative to plastid autofluorescence by calculating the green to red pixel ratio between ten transfected cells with reconstructed YFP signal and an average of 20 randomly chosen plastids from each transfected cell and its surrounding cells. For subcellular localization of the VirF–VIP1 complexes, nYFP–VIP1 and free DsRed2 were expressed from a single plasmid by cloning a 35S promoter–nYFP–VIP1–35S terminator cassette into pGDR²² that already contained the *dsred2* gene under a constitutive promoter, resulting in pGDR–nYFP–VIP1, which was then co-bombarded with cYFP–VirF into tobacco leaves.

Protein destabilization assay in yeast

VIP1, VirE2, or VirD2 fused to the C terminus of GFP, as well as free VIP1, were expressed from the galactose-inducible promoter and free VirF was expressed from the methionine repressible promoter. TAT7 yeast cells, carrying the indicated plasmid combinations, were grown overnight at 30 °C in selective minimal medium with glucose (10% final concentration), washed, and diluted to $A_{600} = 0.1$. For induction of expression, cell cultures were supplemented with galactose (10% final concentration) and allowed to grow for 3 h at 30 °C. At this time, the cells were sampled to determine their GFP fluorescence; this signal represented the initial level of the GFP-tagged protein accumulated in the

expressing cells. The remaining cultures was then collected, washed, and transferred to a glucose-containing growth medium, which either lacked methionine for induction of VirF expression or contained methionine for continuing suppression of VirF synthesis. Three hours later, GFP fluorescence in the live cell cultures was determined, and destabilization of the GFP-tagged proteins was estimated from the reduction in GFP fluorescence calculated as per cent of the initial signal level.

For inactivation of Skp1, the temperature-sensitive *skp1-4* cells, which arrest in the cell cycle²⁴, were maintained at the permissive temperature (25 °C) during overnight growth and galactose induction, and shifted to the restrictive temperature (37 °C) for methionine depletion.

Protein destabilization assay in planta

gfp–VIP1 was cloned into pGDR, and unfused *virF* was cloned into PRTL2, producing pGDR–GFP–VIP1 and PRTL2–VirF, respectively. pGDR–GFP–VIP1 alone or mixed with PRTL2–VirF (1:1 mol/mol) was bombarded into tobacco leaves or onion scales, and the expression patterns of GFP and DsRed2 were analysed by confocal microscopy. Cells transformed with pGDR–GFP–VIP1 were identified by the presence of DsRed2 whereas VIP1 degradation in these cells was determined by reduction or disappearance of the GFP fluorescence.

Early T-DNA expression assay in the presence of MG132

To detect specifically T-DNA expressed in plants, but not in *Agrobacterium*, we chose the binary plasmid pBIG–HYG–GUS that carries in its T-DNA an intron-containing *uidA* gene, the transcript of which is spliced only within plant cells. Tobacco leaf disks were placed on a regeneration medium supplemented with MG132 (25 μ M or 50 μ M in 1% DMSO), incubated for 2 h at 25 °C, bombarded with a GFP-expressing plasmid (PRTL2–GFP), and immediately inoculated with acetosyringone-induced culture (final $A_{600} = 1.5$) of *Agrobacterium* strain EHA105 harbouring pBIG–HYG–GUS. The incubation was continued for 2 h, after which the leaf disks were washed, transferred to a fresh regeneration medium, and sampled at 2, 4, 6, 8, 20 and 24 h after inoculation. Expression of the T-DNA and of the bombarded PRTL2–GFP was analysed by RT–PCR as described²⁷, using *uidA*- and *gfp*-specific primers.

Received 23 April; accepted 12 July 2004; doi:10.1038/nature02857.

- Stachel, S. E. & Zambryski, P. C. Generic trans-kingdom sex? *Nature* **340**, 190–191 (1989).
- Vergunst, A. C. *et al.* VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. *Science* **290**, 979–982 (2000).
- Howard, E., Zupan, J., Citovsky, V. & Zambryski, P. C. The VirD2 protein of *A. tumefaciens* contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells. *Cell* **68**, 109–118 (1992).
- Citovsky, V., Zupan, J., Warnick, D. & Zambryski, P. C. Nuclear localization of *Agrobacterium* VirE2 protein in plant cells. *Science* **256**, 1802–1805 (1992).
- Ballas, N. & Citovsky, V. Nuclear localization signal binding protein from *Arabidopsis* mediates nuclear import of *Agrobacterium* VirD2 protein. *Proc. Natl Acad. Sci. USA* **94**, 10723–10728 (1997).
- Tzfira, T., Vaidya, M. & Citovsky, V. VIP1, an *Arabidopsis* protein that interacts with *Agrobacterium* VirE2, is involved in VirE2 nuclear import and *Agrobacterium* infectivity. *EMBO J.* **20**, 3596–3607 (2001).
- Tzfira, T., Li, J., Lacroix, B. & Citovsky, V. *Agrobacterium* T-DNA integration: molecules and models. *Trends Genet.* **20**, 375–383 (2004).
- Gelvin, S. B. *Agrobacterium*-mediated plant transformation: the biology behind the “gene-jockeying” tool. *Microbiol. Mol. Biol. Rev.* **67**, 16–37 (2003).
- Gelvin, S. B. *Agrobacterium* and plant genes involved in T-DNA transfer and integration. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 223–256 (2000).
- Schrammeijer, B. *et al.* Interaction of the virulence protein VirF of *Agrobacterium tumefaciens* with plant homologs of the yeast Skp1 protein. *Curr. Biol.* **11**, 258–262 (2001).
- Bundock, P., den Dulk-Ras, A., Beijersbergen, A. & Hooykaas, P. J. J. Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. *EMBO J.* **14**, 3206–3214 (1995).
- Kunik, T. *et al.* Genetic transformation of HeLa cells by *Agrobacterium*. *Proc. Natl Acad. Sci. USA* **98**, 1871–1876 (2001).
- Zupan, J., Muth, T. R., Draper, O. & Zambryski, P. C. The transfer of DNA from *Agrobacterium tumefaciens* into plants: a feast of fundamental insights. *Plant J.* **23**, 11–28 (2000).
- Tzfira, T. & Citovsky, V. Partners-in-infection: host proteins involved in the transformation of plant cells by *Agrobacterium*. *Trends Cell Biol.* **12**, 121–129 (2002).
- Melchers, L. S. *et al.* Octopine and nopaline strains of *Agrobacterium tumefaciens* differ in virulence; molecular characterization of the *virF* locus. *Plant Mol. Biol.* **14**, 249–259 (1990).
- Jarchow, E., Grimslay, N. H. & Hohn, B. *virF*, the host range-determining virulence gene of *Agrobacterium tumefaciens*, affects T-DNA transfer to *Zea mays*. *Proc. Natl Acad. Sci. USA* **88**, 10426–10430 (1991).
- Regensburg-Tuink, A. J. & Hooykaas, P. J. J. Transgenic *N. glauca* plants expressing bacterial virulence gene *virF* are converted into hosts for nopaline strains of *A. tumefaciens*. *Nature* **363**, 69–71 (1993).
- del Pozo, J. C. & Estelle, M. F-box proteins and protein degradation: an emerging theme in cellular regulation. *Plant Mol. Biol.* **44**, 123–128 (2000).
- Deng, W. *et al.* VirE1 is a specific molecular chaperone for the exported single-stranded-DNA-binding protein VirE2 in *Agrobacterium*. *Mol. Microbiol.* **31**, 1795–1807 (1999).
- Hu, C. D., Chinenov, Y. & Kerppola, T. K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **9**, 789–798 (2002).
- Dommissie, E. M., Leung, D. W. M., Shaw, M. L. & Conner, A. J. Onion is a monocotyledonous host for *Agrobacterium*. *Plant Sci.* **69**, 249–257 (1990).
- Goodin, M. M., Dietzgen, R. G., Schichnes, D., Ruzin, S. & Jackson, A. O. pGDR vectors: versatile tools for the expression of green and red fluorescent protein fusions in agroinfiltrated plant leaves. *Plant J.* **31**, 375–383 (2002).

23. Gray, W. M., Kepinski, S., Rouse, D., Leyser, O. & Estelle, M. Auxin regulates SCF^{TR1}-dependent degradation of AUX/IAA proteins. *Nature* **414**, 271–276 (2001).
24. Connelly, C. & Hieter, P. Budding yeast *SKP1* encodes an evolutionarily conserved kinetochore protein required for cell cycle progression. *Cell* **86**, 275–285 (1996).
25. Tzfira, T., Vaidya, M. & Citovsky, V. Increasing plant susceptibility to *Agrobacterium* infection by overexpression of the *Arabidopsis* *VIP1* gene. *Proc. Natl Acad. Sci. USA* **99**, 10435–10440 (2002).
26. Lee, D. H. & Goldberg, A. L. Proteasome inhibitors: valuable new tools for cell biologists. *Trends Cell Biol.* **8**, 397–403 (1998).
27. Narasimhulu, S. B., Deng, X.-B., Sarria, R. & Gelvin, S. B. Early transcription of *Agrobacterium* T-DNA genes in tobacco and maize. *Plant Cell* **8**, 873–886 (1996).
28. Otten, L. *et al.* Restoration of virulence of *vir* region mutants of *A. tumefaciens* strain B6S3 by coinfection with normal and mutant *Agrobacterium* strains. *Mol. Gen. Genet.* **195**, 159–163 (1984).
29. Schrammeijer, B., Hemelaar, J. & Hooikaas, P. J. The presence and characterization of a *virF* gene on *Agrobacterium vitis* Ti plasmids. *Mol. Plant Microbe Interact.* **11**, 429–433 (1998).
30. Nagai, H. & Roy, C. R. Show me the substrates: modulation of host cell function by type IV secretion systems. *Cell. Microbiol.* **5**, 373–383 (2003).

Acknowledgements We thank S. Gelvin for VirF and *Agrobacterium* strains, T. Durfee for ASK1, and M. Goodin for pGDR. We are also grateful to R. Sternglanz and A. Neiman for their suggestions and discussion. The work in our laboratory is supported by grants from the National Institutes of Health, National Science Foundation, US Department of Agriculture, US–Israel Bi-national Science Foundation (BSF), and US–Israel Bi-national Agricultural Research and Development Fund (BARD) to V.C., and by grants from BARD and Human Frontiers Science Program to T.T.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.T. (ttzfira@ms.cc.sunysb.edu).

Centrosomes direct cell polarity independently of microtubule assembly in *C. elegans* embryos

Carrie R. Cowan & Anthony A. Hyman

Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, Dresden 01307, Germany

Polarity establishment requires a symmetry-breaking event, resulting in an axis along which determinants are segregated. In *Caenorhabditis elegans*, oocytes are apolar and are triggered to polarize rapidly along one axis after fertilization. The establishment of this first polarity axis is revealed by the asymmetric distribution of PAR proteins and cortical activity in the one-celled embryo. Current evidence suggests that the centrosome–pronucleus complex contributed by the sperm is involved in defining the polarization axis^{1–6}. Here we directly assess the contribution of the centrosome to polarity establishment by laser ablating the centrosome before and during polarization. We find that the centrosome is required to initiate polarity but not to maintain it. Initiation of polarity coincides with the proximity of the centrosome to the cortex and the assembly of pericentriolar material on the immature sperm centrosome. Depletion of microtubules or the microtubule nucleator γ -tubulin did not affect polarity establishment. These results demonstrate that the centrosome provides an initiating signal that polarizes *C. elegans* embryos and indicate that this signalling event might be independent of the role of the centrosome as a microtubule nucleator.

After fertilization, the PAR proteins are uniformly distributed in the *C. elegans* embryo^{5,7–9}. Coincident with the completion of female meiosis, about 30 min after fertilization, shallow ingressions ('ruffles') become visible throughout the cortex. On perception of a polarization signal a few minutes later, the PAR proteins segregate

into their anterior (PAR-3 and PAR-6)^{5,8,9} and posterior (PAR-1 and PAR-2)^{5,7,10} domains. Concomitantly, contractile polarity is established⁵: the anterior cortex continues to ruffle while cortical activity in the posterior ceases, creating a smooth domain. The male pronucleus and microtubule asters lie adjacent to the smooth, posterior PAR-2 domain^{5,6}. To investigate when centrosome–cortex proximity occurs relative to the initiation of polarity, centrosome position was monitored with a green fluorescent protein (GFP)-tagged centrosome marker, SPD-2 (ref. 11), and polarity was assessed both by GFP–PAR-2 (Fig. 1a; Supplementary Movie 1) and by cortical ruffling. The initiation of cortical polarity was evident as a progressive cessation of ruffling, spreading from a focus on the cortex, and within 2 min GFP–PAR-2 was visible on the expanding smooth cortex. Before the onset of polarization, the centrosome–cortex distances were variable (Fig. 1b; the range at –3:20 (3 min 20 s before the onset of posterior smoothing) was 0–9 μ m; $n = 9$). However, we consistently observed that the centrosome was closest to the cortex when cortical polarity was initiated (Fig. 1b; range at 0:00 was 0–4 μ m; $n = 14$). Thus, polarity initiation coincided with centrosome–cortex proximity.

As a direct test of the role of the centrosome in polarity establishment, centrosomes were located during female meiosis with the use of GFP–SPD-2 fluorescence and the labelled centrosomes were ablated with a pulsed ultraviolet-laser microbeam. Polarity was assessed by time-lapse imaging of GFP–PAR-2 or contractile polarity. Successful ablations were judged by the absence of a GFP–SPD-2 structure throughout the recording period (minimally 30 min after ablation) and a failure of the fast phase of pronuclear migration, a phenotype associated with centrosome defects. In embryos in which the centrosome was successfully ablated before the onset of polarity, no asymmetric PAR-2 localization was observed ($n = 30$; Fig. 2a; Supplementary Movies 2 and 3) and ruffling continued throughout the entire cortex (Supplementary

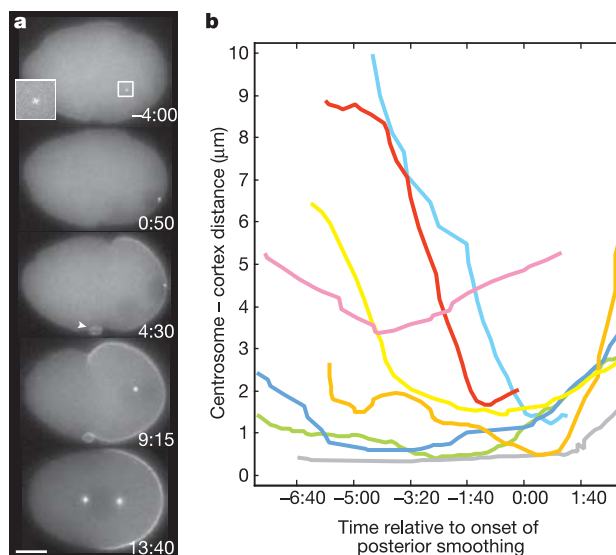


Figure 1 The centrosome lies adjacent to the cortex at the time of polarity initiation. **a**, Time lapse images of a GFP–PAR-2; GFP–SPD-2 embryo (Supplementary Movie 1). The GFP–SPD-2-labelled centrosome is visible as a bright cytoplasmic dot (the second centrosome is out of the focal plane); GFP–PAR-2 labels the posterior cortex. The small ring of GFP–PAR-2 (arrowhead, 4:30) corresponds to the polar body cortex. The embryo posterior is to the right. Scale bar, 10 μ m. **b**, Centrosome–cortex distances during polarity establishment. Solid lines indicate the distance (μ m) from the centrosome to the nearest point on the cortex over time. PAR-2 polarity was established at about 2:00 (see Methods). Times were standardized to the onset of posterior smoothing, indicating polarity establishment. Only a subset of experiments is shown.

Movie 2). In the experiments in which GFP-SPD-2 reappeared after less than 5 min, indicating that the centrosome had probably been photobleached rather than destroyed, normal PAR-2 polarity was established. Similarly, if the nucleus or cytoplasm in the vicinity of the centrosome, or a cortical region away from the centrosome, was

ablated, asymmetric PAR-2 localization occurred normally ($n = 19$; Supplementary Movie 4), indicating that only centrosome ablation inhibits the establishment of polarity. We conclude that the centrosome, or a closely associated cellular component, is required to establish polarity in the one-cell *C. elegans* embryo.

The centrosome could be required continuously during polarity establishment or could provide a transient symmetry-breaking signal. To distinguish between these ideas, we laser ablated the centrosome at different times during polarity establishment. If embryos had initiated but not yet completed polarization, the GFP-PAR-2 domain expanded in the absence of the centrosome (Fig. 2b; Supplementary Movie 5). If the centrosome was ablated after polarity establishment, polarity was maintained (Fig. 2c; Supplementary Movie 6). We therefore conclude that the centrosome is required to initiate polarity but not to propagate or maintain it.

Although polarity maintenance occurred in the absence of centrosomes, the PAR-2 domain was smaller than in control embryos. In control ablation embryos, PAR-2 expanded to $53 \pm 10\%$ egg length ($n = 8$). If the centrosomes were ablated in the early stages of PAR-2 spreading, PAR-2 expanded to only $37 \pm 7\%$ egg length ($n = 14$). If the centrosomes were ablated after the domain had formed (PAR-2 extended to about 50% of egg length), PAR-2 seemed to regress ($n = 5$). All embryos showed a narrow range of final GFP-PAR-2 extension after centrosome ablation (Fig. 2d, black dots), despite having various sizes of initial posterior domains (Fig. 2d, grey bars). We therefore conclude that expansion of the PAR-2 domain is largely independent of the centrosome but that full extension of the domain requires a centrosome-dependent process. Taken together, our experiments indicate that the centrosomes might provide a transient signal that determines the polarity of the *C. elegans* embryo, but propagation of this polarity can occur without centrosomes.

The centrosome contributed by the sperm is immature and is therefore incapable of microtubule nucleation. About 30 min after fertilization, the centrosome accumulates pericentriolar material¹² (PCM) and initiates microtubule nucleation¹³. To relate the state of centrosome assembly to the onset of polarity, we tracked the fluorescence intensity of GFP-labelled SPD-5, a centrosomal protein that is required for the accumulation of the known PCM components¹. SPD-5 appeared on the centrosome roughly simultaneously with polarity initiation (Fig. 3a, red line). At about the same time, the microtubule nucleator γ -tubulin^{12,14} accumulated on the centrosome (Fig. 3a, blue line). We therefore conclude that polarity initiation occurs concomitantly with the assembly of the PCM at the centrosome.

To investigate the role of PCM assembly in polarity establishment, we first analysed the role of SPD-5, which was previously implicated in PAR protein polarity¹. We tracked the formation of the GFP-PAR-2 domain over time (Fig. 3b–f) and plotted the boundary (solid green lines) and domain extent (dotted green lines) of cortical PAR-2 and the movement of the pronuclei (blue dots). In control embryos, cortical PAR-2 spread from the posterior pole to the embryo middle during a period of roughly 7 min ($n = 44$; Fig. 3b). In embryos depleted of SPD-5 by RNA-mediated interference (*spd-5(RNAi)*), GFP-PAR-2 was found only in the cytoplasm ($n = 15$; Fig. 3c) and ruffling occurred throughout the entire embryo cortex (Supplementary Fig. 1C) during an equivalent time interval. Similar results were obtained after depletion of SPD-2 ($n = 27$; Fig. 3d; Supplementary Fig. 1D), a protein required for the recruitment of SPD-5 to the centrosome^{1,11,15}, as shown previously². Thus we conclude that the depletion of PCM components results in the same polarity phenotype as centrosome ablation, which is consistent with previous data^{1,2}. We tracked the proximity of the centrosome to the cortex in *spd-5(RNAi)* and *spd-2(RNAi)* embryos and found that centrosome–cortex distances were within the range of control embryos (Supplementary Fig. 2), indicating that both

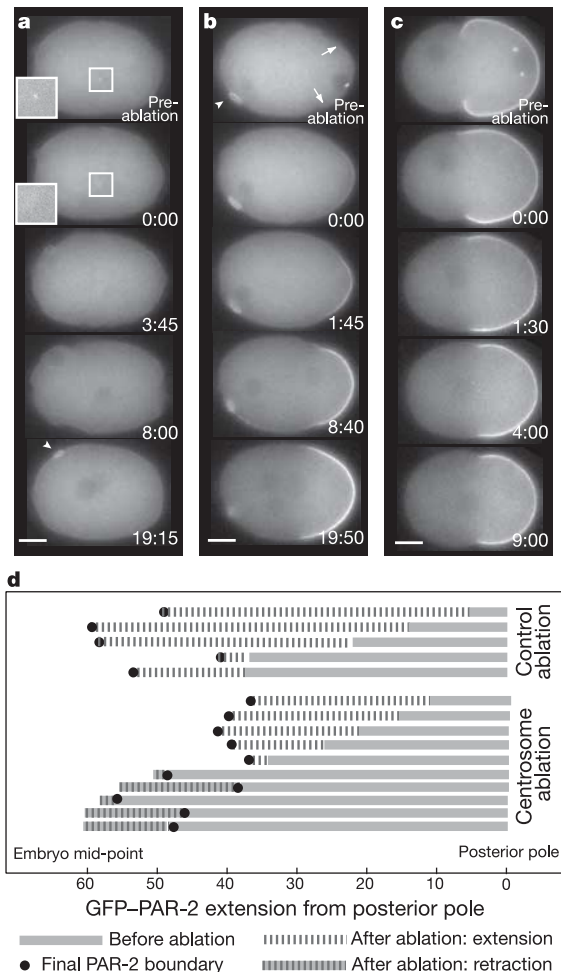


Figure 2 Centrosome ablation prevents polarity establishment but not the propagation and maintenance of the posterior domain. **a–c**, Time-lapse images of GFP-PAR-2 polarity after centrosome ablation. The first frame ('pre-ablation') in each series immediately precedes the ablation; the time elapsed after centrosome ablation is indicated in minutes:seconds. The embryo posterior is to the right. Scale bars, 10 μ m. **a**, Centrosome ablation before polarity establishment (Supplementary Movies 2 and 3). Embryos in which the centrosomes were not adjacent to the cortex were chosen for ablation, to eliminate the possibility of damaging the cortex. Either one or two closely spaced GFP-SPD-2 dots were visible. **b**, Centrosome ablation during polarity establishment (Supplementary Movie 5). Ablations were performed on embryos in which the centrosomes (two closely spaced GFP-SPD-2 dots) were near the cortex and the posterior domain (judged by ruffling and GFP-PAR-2) occupied less than half the embryo length. Arrows indicate the boundaries of the 'pre-ablation' posterior domain; GFP-PAR-2 is present only weakly although ruffling has ceased. In **a** and **b** the small ring of GFP-PAR-2 at the anterior (arrowhead) corresponds to the polar body cortex. **c**, Centrosome ablation after polarity establishment (Supplementary Movie 6). Polarity was considered to be established when the posterior domain occupied half the embryo length. The centrosomes had moved away from the cortex and separated around the nuclear envelope. The two centrosomes were ablated individually. **d**, Relation between pre-ablation and post-ablation PAR-2 domain extension. Bars represent the percentage of the anterior–posterior axis of the embryo occupied by cortical PAR-2. Control ablations represent embryos in which the cortex rather than the centrosome was ablated (Supplementary Movie 7). Only a subset of experiments is shown.

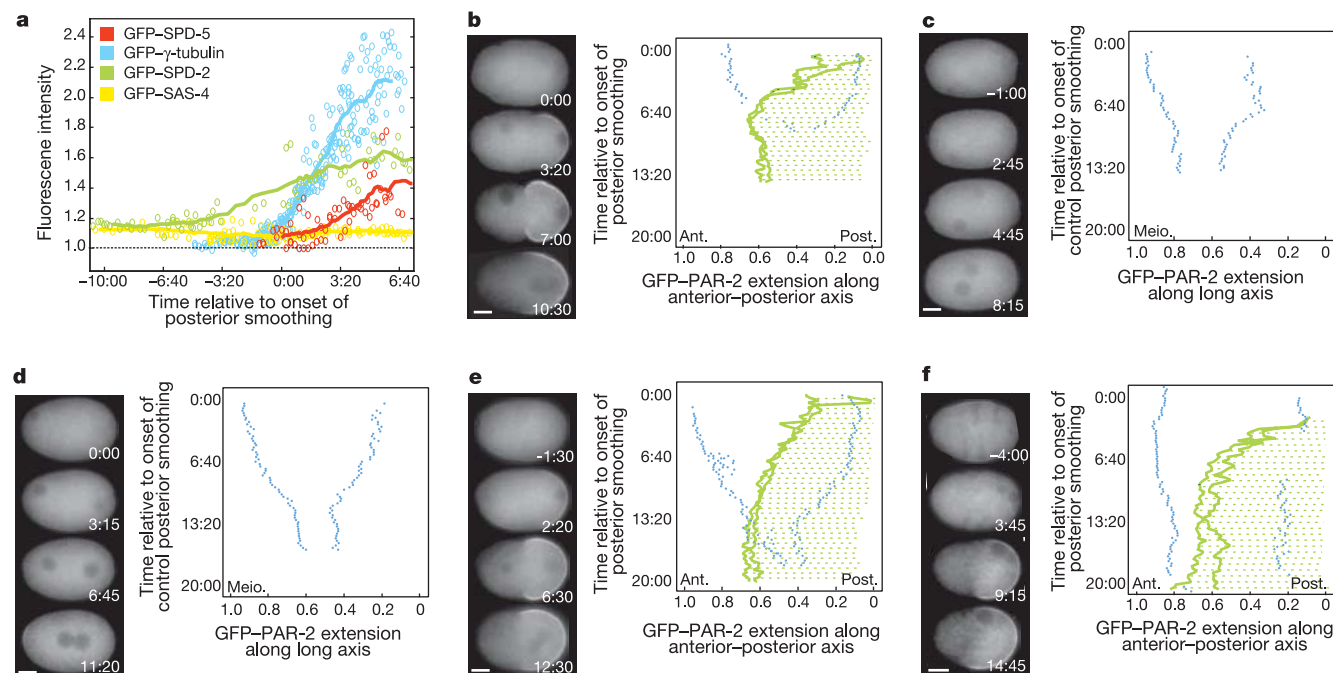


Figure 3 Polarity initiation requires a concomitant assembly of the pericentriolar material. **a**, Kinetics of centrosome assembly relative to polarity establishment. Centrosomal fluorescence is represented as the ratio of cytoplasmic background (1.0 represents no signal). SAS-4, a sperm-provided centriole protein, is shown for comparison. The data represent a minimum of five embryos per centrosome marker. **b–f**, Time-lapse images and kymographs (see Methods) of GFP-PAR-2 polarity in control (**b**), *spd-5(RNAi)* (**c**), *spd-2(RNAi)* (**d**), γ -tubulin(*RNAi*) (**e**) and α -/ β -tubulin(*RNAi*) (**f**) embryos (similar results were obtained for nocodazole treatment and β -tubulin(*RNAi*)/nocodazole; Supplementary Fig. 4). Solid green lines, PAR-2 boundaries; dotted green lines, cortical PAR-2; blue dots, pronuclei. Embryo posterior is to the right. Anterior (ant.) and posterior (post.), or the meiotic pole (meio.) in cases where polarity was not

established, are indicated on the graphs; all extensions are shown as fractions of the embryo length. Scale bars, 10 μ m. In **a**, **b**, **e** and **f**, times are standardized to the onset of posterior smoothing; in **c** and **d**, times were assigned on the basis of a similar cell cycle stage to that in control embryos at the time of polarization, judged by the size of the male pronucleus and the time elapsed relative to meiosis II. Although *spd-5(RNAi)* and *spd-2(RNAi)* embryos showed no asymmetric PAR-2 distribution during the period of control polarity establishment, some embryos developed small polar patches of PAR-2 at later times (*spd-5(RNAi)*, 11:30 $\pm$ 3:40 ($n = 10$); *spd-2(RNAi)*, 11:00 $\pm$ 2:20 ($n = 4$)), which is consistent with the polarity defects described for both *spd-5* and *spd-2* mutants^{1,2}.

SPD-2 and SPD-5 might be required for the polarizing signal.

The depletion of centrosome components thought to be required directly for efficient microtubule nucleation did not affect polarity. Normal polarity was established after the depletion of γ -tubulin ($n = 36$; Fig. 3e; Supplementary Fig. 1E), although no centrosomal microtubules were visible at the time of polarity initiation¹² (Fig. 4a; Supplementary Fig. 3). Similar results were seen after the depletion of other components of the γ -tubulin complex¹² (data not shown). Conversely, the centrosome in *spd-2(RNAi)* embryos nucleated wild-type levels of microtubules at the time of polarity initiation (Fig. 4a; Supplementary Fig. 3), but polarity was not established (Fig. 3d; Supplementary Fig. 1D). There is therefore no obvious correlation between centrosomal microtubule nucleation and polarity establishment.

To further assess the role of microtubules in polarity establishment, we depleted microtubules by using α -/ β -tubulin(*RNAi*), treatment with nocodazole, or a combination of β -tubulin(*RNAi*)/nocodazole treatment. Microtubule depletion had no effect on the asymmetric localization of GFP-PAR-2 ($n = 27$; Figs 3f and 4) or the polarization of ruffling activity (Supplementary Fig. 1F; Supplementary Fig. 4). Thus polarity establishment in *C. elegans* can occur without a detectable population of microtubules ($n = 3$; Fig. 4b). It seems that the centrosome performs separable functions in the early *C. elegans* embryo, namely microtubule nucleation and polarity establishment.

Previous work indicated that microtubules were sufficient to establish a PAR-2 domain in *C. elegans* embryos during meiosis. Meiosis-arrested embryos exhibited a small cortical PAR-2 patch

overlying the meiotic spindle, and formation of this domain was partly inhibited by microtubule depolymerization. This meiotic PAR-2 domain appears transiently in wild-type embryos and seems to form the polar body cortex (Figs 1 and 2). We have found that neither SPD-2 nor SPD-5 is required for the formation of the meiotic PAR-2 domain (C.R.C. and A.A.H., unpublished observations), whereas both are required for embryonic polarity establishment^{1,2} (Fig. 3; Supplementary Fig. 1). The mechanisms by which PAR-2 is spatially directed to the cortex therefore differ during meiosis and the subsequent establishment of polarity, perhaps reflecting differences in the underlying mechanisms by which the cortical domains are formed.

On the basis of our experiments, we propose that the centrosome becomes positioned near the embryo cortex simultaneously with the assembly of the PCM about 30 min after fertilization. A signal from the centrosome, which is independent of microtubules, activates a local change in the cortex that, once initiated, propagates independently of the centrosome. The physical change induced in the cortex would result in the cessation of ruffling and allow PAR-2 to localize. Our experiments suggest that both SPD-5 and SPD-2 are required directly for the polarity signal produced by the centrosome.

The proposed centrosome signalling event in the *C. elegans* embryo is reminiscent of microtubule-independent centrosome-mediated reorganization of the cell cortex during blastoderm cellularization in *Drosophila*¹⁶, asymmetric division of *Drosophila* neuroblast precursors¹⁷, and cytokinesis completion in some tissue culture cells¹⁸. Centrosome-mediated cortical changes are therefore

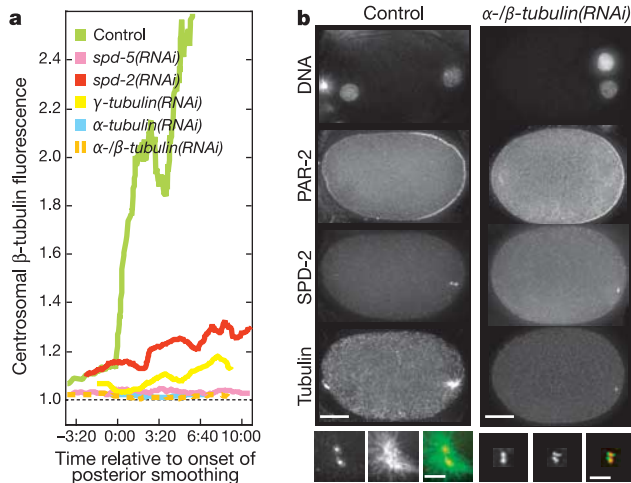


Figure 4 Polarity establishment and microtubule assembly are separable centrosomal functions. **a**, Centrosomal GFP-β-tubulin accumulation after depletion of SPD-5, SPD-2, γ-tubulin, α-tubulin and α-/β-tubulin. Fluorescence intensity is represented as the ratio of cytoplasmic background (1.0 represents no signal). Times are standardized to the onset of posterior smoothing, except for *spd-5(RNAi)* and *spd-2(RNAi)* embryos, for which times were assigned on the basis of a similar cell cycle stage to that in control embryos at the time of polarization, judged by the size of the male pronucleus and the time elapsed relative to meiosis II. The data represent at least five embryos per RNAi condition. **b**, Microtubules and PAR-2 polarity in control and *α-/β-tubulin(RNAi)* embryos. Tubulin images are projections of deconvolved z-sections representing the entire cell; PAR-2 and DNA images are projections comprising about one-third of the cell; SPD-2 images are projections encompassing the centrosomal signals. The embryo posterior is to the right. Scale bars, 10 μm. Magnifications correspond to the centrosomal region: left, SPD-2; middle, tubulin; right, merge (red, SPD-2; green, tubulin). Scale bars, 2.5 μm.

likely to occur in many cellular contexts, indicating that they might share an evolutionarily conserved mechanism. □

Methods

Worm strains and time-lapse imaging

All worm strains were maintained at 16 °C and shifted to 25 °C for 20–30 h before recording. Recording of control and RNAi embryos was performed at 20–23 °C; all ablations were performed at 18–20 °C. Embryos were dissected and mounted¹⁹ in a solution containing 0.1 M NaCl and 4% sucrose, with or without 2% agarose. For PAR-2 and centrosomal marker imaging, wide-field GFP and differential interference contrast (DIC) movies were recorded at 15-s intervals, as described previously²⁰; GFP-β-tubulin movies were acquired with the use of spinning-disk confocal imaging at 10-s intervals, as described previously²⁰. The following strains were used: N2 (wild type), AZ244 (GFP-β-tubulin)²¹, JH1380 (GFP-PAR-2)⁶, TH26 (GFP-SAS-4)²², TH27 (GFP-γ-tubulin)¹², TH40 (GFP-SPD-5), TH42 (GFP-SPD-2)¹¹ and TH49 (GFP-PAR-2; GFP-SPD-2). Amino-terminal GFP-tagged SPD-5 is expressed under the *pie-1* promoter, and transgenic worms were created by microparticle bombardment, as described previously²¹.

Data analysis

Data analysis was performed with Matlab (The Math Works, Inc.) and Metamorph (Universal Imaging). Results are given as means ± standard deviations.

Centrosome-cortex distances

Distances were calculated from embryos in which the centrosome remained roughly in the middle of the embryo z-axis during the recording interval, to simplify the measurement by restricting it to two dimensions. The distance from the centrosome to the nearest point on the embryo cortex (judged as the edge of cytoplasmic GFP) was measured at each time point. Data were smoothed with a running average of five time points. The onset of cortical smoothing was judged from the corresponding DIC images as the first instance of non-ruffling that was propagated away from its initiation site. Establishment of GFP-PAR-2 asymmetry was determined mathematically by comparing maximum intensity values for a 16 pixel × 16 pixel region of the anterior and posterior cortices. A posterior:anterior PAR-2 ratio greater than 1.5 was considered asymmetric. Centrosome-cortex distances behaved similarly in GFP-SPD-2, GFP-PAR-2; GFP-SPD-2 and GFP-SAS-4 embryos (data not shown).

Laser ablations

Laser ablations were performed with the experimental set-up described previously²³, with the following changes. Laser power (before entry into the microscope) was attenuated to less than 5%. About 200 laser pulses at 1,000 MHz were delivered along a 1-μm line through the centrosome. Wide-field GFP images were recorded as described²³.

Centrosome ablations before polarity establishment were performed only in embryos in which the centrosome was at least about 3 μm away from the cortex and roughly in the middle of the z-axis of the embryo, to avoid damage to the cortex. Control ablations in pre-polarity embryos were performed by targeting the nucleus (*n* = 4) or cytoplasm (*n* = 8) within about 5 μm of the centrosome (and away from the cortex) or the cortex away from the centrosome (*n* = 7) and ablating the region under the same conditions as for centrosome ablations. Centrosome ablation during polarity establishment necessitated that we ablate the centrosomes within about 2 μm of the cortex; controls for such ablations were therefore performed by targeting a region of the cortex (within about 2 μm of the outermost extent of GFP-PAR-2) away from the position of the centrosomes (about 5 μm) and ablating the region as for centrosome ablations (Supplementary Movie 7).

PAR-2 domain extension

PAR-2 domain sizes, from posterior pole to anterior PAR-2 boundary, were measured as a distance along the long axis of the embryo. The distances were standardized to the length of the embryo (anterior-posterior axis distance) and are expressed as percentages of embryo length; 0% indicates the posterior pole.

PCM and β-tubulin accumulation

The integrated fluorescence intensity of a circle with a radius of three pixels (roughly the size of an early prophase centrosome) was calculated for both the centrosome and a nearby region of cytoplasm (background) at each time point. Times were standardized to the onset of posterior smoothing. The ratio of centrosomal fluorescence to background over time was plotted for a minimum of five individual embryos for each centrosomal marker (Fig. 3a, individual circles) or each RNAi condition (Fig. 4, data not shown); the combined data sets were subjected to a running average (window size 5, lines in Fig. 3a; window size 25, Fig. 4a). In *α-tubulin(RNAi)*, *α-/β-tubulin(RNAi)* and *spd-5(RNAi)* embryos there was no evident centrosomal GFP-β-tubulin and we therefore calculated the fluorescence intensity for bright GFP-β-tubulin foci in the region of the male pronucleus.

RNAi

RNAi experiments were performed as described²⁰. The primers used for the production of double-stranded RNA *in vitro* are listed in Supplementary Table S1. Worms were grown for 22–25 h (depending on the individual double-stranded RNA) at 25 °C after injection.

PAR-2 tracking

We manually generated kymographs of PAR-2 boundaries from time-lapse images. The *xyt* position of the anterior edge of GFP-PAR-2 was tracked manually. The maximum embryo length (distance from anterior to posterior) from the recording was standardized to 1.0 and all other values are given as fractions of this maximum embryo length. The anterior boundaries of PAR-2 (solid lines) were projected onto the anterior-posterior axis (or the long axis if polarity was not established), displayed as the x-axis of the graphs; the cortical PAR-2 domain is indicated by dotted lines. The position of PAR-2 was plotted for each time point, represented along the y-axis. Migration of the pronucleus was plotted by using the same process as that described for PAR-2 boundaries; *xyt* positions of the nuclei were determined by the exclusion of cytoplasmic GFP from the images.

Immunofluorescence

Immunofluorescence was performed as described¹⁹ with the use of DM1α (1:100; Sigma) and PAR-2a²⁴ (1:100) antibodies were visualized with fluorescent-conjugated secondary antibodies (Jackson Immunochemicals), and a directly labelled SPD-2 antibody¹¹ was used to detect centrosomes. Wide-field images were acquired and deconvolved as described²⁰.

Received 17 March; accepted 7 July 2004; doi:10.1038/nature02825.

- Hamill, D. R., Severson, A. F., Carter, J. C. & Bowerman, B. Centrosome maturation and mitotic spindle assembly in *C. elegans* require SPD-5, a protein with multiple coiled-coil domains. *Dev. Cell* **3**, 673–684 (2002).
- O'Connell, K. F., Maxwell, K. N. & White, J. G. The *spd-2* gene is required for polarization of the anteroposterior axis and formation of the sperm asters in the *Caenorhabditis elegans* zygote. *Dev. Biol.* **222**, 55–70 (2000).
- Goldstein, B. & Hird, S. N. Specification of the anteroposterior axis in *Caenorhabditis elegans*. *Development* **122**, 1467–1474 (1996).
- Sadler, P. L. & Shakes, D. C. Anucleate *Caenorhabditis elegans* sperm can crawl, fertilize oocytes and direct anterior-posterior polarization of the 1-cell embryo. *Development* **127**, 355–366 (2000).
- Cuenca, A. A., Schetter, A., Aceto, D., Kemphues, K. & Seydoux, G. Polarization of the *C. elegans* zygote proceeds via distinct establishment and maintenance phases. *Development* **130**, 1255–1265 (2003).
- Wallenfang, M. R. & Seydoux, G. Polarization of the anterior-posterior axis of *C. elegans* is a microtubule-directed process. *Nature* **408**, 89–92 (2000).
- Boyd, L., Guo, S., Levitan, D., Stinchcomb, D. T. & Kemphues, K. J. PAR-2 is asymmetrically distributed and promotes association of P granules and PAR-1 with the cortex in *C. elegans* embryos. *Development* **122**, 3075–3084 (1996).
- Hung, T. J. & Kemphues, K. J. PAR-6 is a conserved PDZ domain-containing protein that colocalizes with PAR-3 in *Caenorhabditis elegans* embryos. *Development* **126**, 127–135 (1999).
- Etemad-Moghadam, B., Guo, S. & Kemphues, K. J. Asymmetrically distributed PAR-3 protein contributes to cell polarity and spindle alignment in early *C. elegans* embryos. *Cell* **83**, 743–752 (1995).
- Guo, S. & Kemphues, K. J. *par-1*, a gene required for establishing polarity in *C. elegans* embryos,

- encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611–620 (1995).
11. Pelletier, L. *et al.* The *Caenorhabditis elegans* centrosomal protein SPD-2 is required for both pericentriolar material recruitment and centriole duplication. *Curr. Biol.* **14**, 863–873 (2004).
 12. Hannak, E. *et al.* The kinetically dominant assembly pathway for centrosomal asters in *Caenorhabditis elegans* is gamma-tubulin dependent. *J. Cell Biol.* **157**, 591–602 (2002).
 13. Albertson, D. G. Formation of the first cleavage spindle in nematode embryos. *Dev. Biol.* **101**, 61–72 (1984).
 14. Strome, S. *et al.* Spindle dynamics and the role of gamma-tubulin in early *Caenorhabditis elegans* embryos. *Mol. Biol. Cell* **12**, 1751–1764 (2001).
 15. Kemp, C. A., Kopish, K. R., Zipperlen, P., Ahringer, J. & O'Connell, K. F. Centrosome maturation and duplication in *C. elegans* require the coiled-coil protein SPD-2. *Dev. Cell* **6**, 511–523 (2004).
 16. Stevenson, V. A., Kramer, J., Kuhn, J. & Theurkauf, W. E. Centrosomes and the Scrambled protein coordinate microtubule-independent actin reorganization. *Nature Cell Biol.* **3**, 68–75 (2001).
 17. Berdnik, D. & Knoblich, J. A. *Drosophila* Aurora-A is required for centrosome maturation and actin-dependent asymmetric protein localization during mitosis. *Curr. Biol.* **12**, 640–647 (2002).
 18. Piel, M., Nordberg, J., Euteneuer, U. & Bornens, M. Centrosome-dependent exit of cytokinesis in animal cells. *Science* **291**, 1550–1553 (2001).
 19. Gonczy, P. *et al.* Dissection of cell division processes in the one cell stage *Caenorhabditis elegans* embryo by mutational analysis. *J. Cell Biol.* **144**, 927–946 (1999).
 20. Oegema, K., Desai, A., Rybina, S., Kirkham, M. & Hyman, A. A. Functional analysis of kinetochore assembly in *Caenorhabditis elegans*. *J. Cell Biol.* **153**, 1209–1226 (2001).
 21. Praitis, V., Casey, E., Collar, D. & Austin, J. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157**, 1217–1226 (2001).
 22. Kirkham, M., Muller-Reichert, T., Oegema, K., Grill, S. & Hyman, A. A. SAS-4 is a *C. elegans* centriolar protein that controls centrosome size. *Cell* **112**, 575–587 (2003).
 23. Grill, S. W., Howard, J., Schaffer, E., Stelzer, E. H. & Hyman, A. A. The distribution of active force generators controls mitotic spindle position. *Science* **301**, 518–521 (2003).
 24. Pichler, S. *et al.* OOC-3, a novel putative transmembrane protein required for establishment of cortical domains and spindle orientation in the P(1) blastomere of *C. elegans* embryos. *Development* **127**, 2063–2073 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank: A. Desai, S. Eaton, C. Hoege, K. Oegema, N. Özlü, L. Pelletier, S. Quintin, A. Schlaitz, S. Schonegg, M. Srayko, J. Stear, A. Tudor-Constantinescu and M. van Breugel for comments on the manuscript; S. Grill for advice on laser ablation; A. Pozniakovsky for modified GFP vectors; and G. Seydoux for the gift of JH1380 worms. Some of the worm strains used in this study were obtained from the *Caenorhabditis* Genetics Center, funded by the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.R.C. (cowan@mpi-cbg.de).

Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L

Déborah Bourc'his & Timothy H. Bestor

Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, 701 West 168th Street, New York, New York 10032, USA

Mammalian genomes employ heritable cytosine methylation in the long-term silencing of retrotransposons and genes subject to genomic imprinting and X chromosome inactivation. Little is known of the mechanisms that direct cytosine methylation to specific sequences. Here we show that DNA methyltransferase 3-like (Dnmt3L (ref. 1)) is expressed in testes during a brief perinatal period in the non-dividing precursors of spermatogonial stem cells at a stage where retrotransposons undergo *de novo* methylation. Deletion of the *Dnmt3L* gene prevented the *de novo* methylation of both long-terminal-repeat (LTR) and non-LTR retrotransposons, which were transcribed at high levels in spermatogonia and spermatocytes. Loss of Dnmt3L from early germ cells also caused meiotic failure in spermatocytes, which do not express Dnmt3L. Whereas dispersed repeated sequences were

demethylated in mutant germ cells, tandem repeats in pericentric regions were methylated normally. This result indicates that the Dnmt3L protein might have a function in the *de novo* methylation of dispersed repeated sequences in a premeiotic genome scanning process that occurs in male germ cells at about the time of birth.

Lifelong gene silencing is imposed on target sequences by *de novo* methylation in germ cells and early embryos. The only factors known to be involved in the establishment of methylation patterns in germ cells are Dnmt3A (ref. 2) and Dnmt3L, a protein that lacks the catalytic motifs that characterize the DNA cytosine-5-methyltransferases but is related to the active DNA methyltransferases Dnmt3A and Dnmt3B in framework regions¹. Deletion of *Dnmt3L* does not prevent oogenesis, but the heterozygous offspring of homozygous mutant females die before mid-gestation as a result of biallelic expression of imprinted genes normally methylated and silenced on the allele of maternal origin³. Male mice that lack Dnmt3L are viable but sterile, with a complete absence of germ cells in adult males³.

As shown in Fig. 1, expression of Dnmt3L is first seen in non-dividing prospermatogonia after 12.5 days post coitus (d.p.c.) and is highest at about the time of birth; expression declines rapidly after birth and is extinguished by 6 days post partum (d.p.p.), when most spermatogonia have differentiated into dividing spermatogonial stem cells. Other data show that the decline in Dnmt3L expression over this period is more than 200-fold⁴. Dnmt3L shows striking sexual dimorphism in expression patterns; it is present in females only in growing oocytes, which are arrested in the diplotene stage of meiosis I, but in males it is restricted to diploid prospermatogonia, which differentiate into spermatogonia and undergo many mitotic divisions before entry into meiosis.

The germ cell population of testes lacking Dnmt3L was normal at birth, but only spermatogonia and leptotene and zygotene spermatocytes were seen in testes of young adult males; progressive loss of spermatogonia caused complete azoospermia in older mutant animals. Examination of meiotic chromosome spreads stained with antibodies against synaptonemal complex proteins 1 and 3 (Scp1 and Scp3 (ref. 5)) showed widespread nonhomologous synapsis with branching and anastomosing arrays of synaptonemal complex proteins in most leptotene and zygotene spermatocytes (compare control in Fig. 2a to mutant in Fig. 2b). Some leptotene spermatocytes had complexes of synaptonemal proteins in the form of interlocked rings and other highly abnormal structures (Fig. 2c, d). All spermatocytes showed asynapsis or abnormal synapsis, and none progressed to the full pachytene stage, as assessed by the

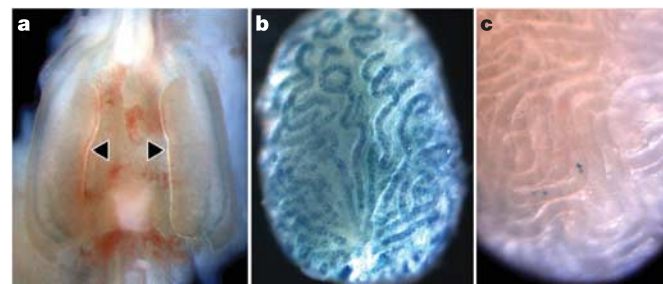


Figure 1 Expression of Dnmt3L in perinatal male germ cells. Dnmt3L transcription was revealed by staining whole testes for β -galactosidase in animals heterozygous for the deletion- β -geo insertion allele of *Dnmt3L* described in ref. 3. **a**, Lack of *Dnmt3L* expression in testes (black arrows) at 12.5 d.p.c. **b**, High-level expression of *Dnmt3L* in germ cells at 2 d.p.p.; the β -galactosidase staining visible in the seminiferous tubules was confirmed to be in prospermatogonia by examination of sectioned specimens. **c**, Loss of *Dnmt3L* expression in germ cells in seminiferous tubules of 6 d.p.p. testis.

lack of pachytene-specific histone H1t and of sex bodies after staining with antibodies against histone γ -H2AX (data not shown). The meiotic abnormalities do not indicate a role for Dnmt3L during normal chromosome synapsis, because the protein is not expressed in meiotic cells, and at the end of the reproductive life span normal meiotic male germ cells can be more than 2 years distant from exposure to Dnmt3L.

The loss of post-zygotene spermatocytes from Dnmt3L-deficient males is likely to arise through the activation of an apoptotic checkpoint triggered before pachynema by non-synapsed chromosomal regions⁶, but the death of spermatogonia in adult males implies that a heritable epigenetic defect is incurred in the perinatal prospermatogonia while under Dnmt3L deprivation and is transmitted by mitotic inheritance during the division of spermatogonia. To test for DNA methylation abnormalities that might underlie the heritable defect, germ cells were isolated from dissociated testes by fluorescence-activated cell sorting after labelling with antibodies against germ cell nuclear antigen-1 (GCNA1 (ref. 7)), and DNA was purified and tested for resistance to the methylation-sensitive restriction endonuclease *HpaII*. As shown in Fig. 3a, DNA of mutant germ cells was less methylated than DNA from wild-type germ cells or mutant somatic cells from testis. The DNA blot hybridization of Fig. 3b shows that the promoter regions of LINE-1 retrotransposons were largely demethylated, as were LTR sequences of retrotransposons of the intracisternal A particle (IAP) class (Fig. 3c). These findings are consistent with the overall demethylation shown in Fig. 3a, because retrotransposons and their remnants contain the large majority of the 5-methylcytosine present

in the mammalian genome⁸. These data show that dispersed repeats within the euchromatic compartment of the genome require Dnmt3L for methylation, but the pericentric tandem repeat DNAs known as major and minor satellites are not dependent on Dnmt3L

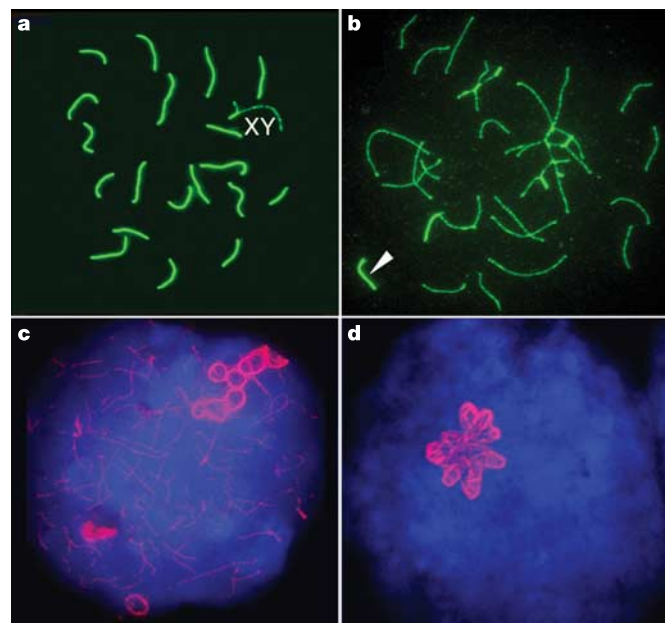


Figure 2 Meiotic catastrophe in spermatocytes derived from Dnmt3L-deficient prospermatogonia. **a**, Normal meiotic synapsis in wild-type spermatocytes. Note the formation of complete linear synaptonemal complexes except for the XY chromosome pair, in which synapsis is restricted to the pseudoautosomal regions. **b**, Branching and anastomosing synaptonemal complexes in Dnmt3L-deficient spermatocytes. Nearly all the chromosomal regions are unpaired or engaged in non-homologous synapsis; a single apparently normal synaptonemal complex is indicated by the white arrowhead. **c**, **d**, Formation of highly aberrant complexes of synaptonemal proteins in the form of interlocked rings (**c**) and complex three-dimensional structures (**d**) in Dnmt3L mutant spermatocytes. Similar staining patterns were seen after labelling with antibodies against Scp3 or a combination of Scp1 and Scp3 antibodies. Meiotic spreads were prepared as described²². Synaptonemal complex proteins are stained in green in **a** and **b**, and in red in **c** and **d**.

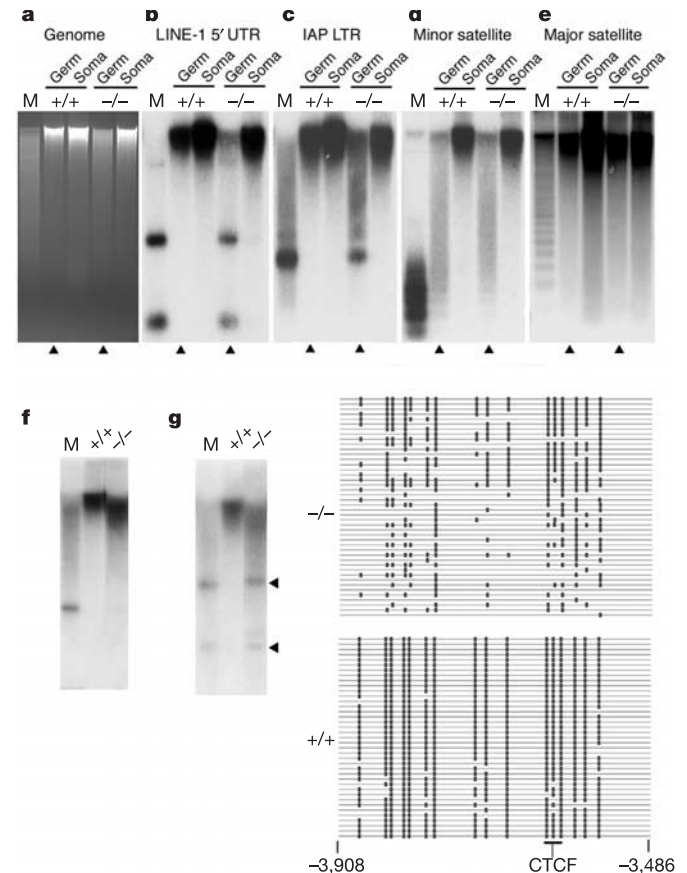


Figure 3 Abnormal genome methylation in Dnmt3L-deficient male germ cells. Germ cells from 17 d.p.p. testes were purified by fluorescence-activated cell sorting after being stained for GCNA1. DNA was digested with the methylation-sensitive restriction endonuclease *HpaII* before DNA blot analysis with the indicated probes. Lanes headed M contained DNA that had been cleaved with *MspI*, a methylation-insensitive isoschizomer of *HpaII*. Soma, somatic cells from testis. **a**, Demethylation of mutant germ-cell genomes revealed by staining DNA with ethidium bromide; arrowheads at the bottom indicate wild-type and Dnmt3L mutant germ cells. **b**, **c**, Demethylation of the promoter region of A-type LINE-1 elements (**b**) and of LTR sequences of IAP elements (**c**) in the absence of Dnmt3L. **d**, **e**, Equal methylation of minor (**d**) and major (**e**) satellites in Dnmt3L-deficient germ cells; satellite DNA in normal male germ cells is not fully methylated^{23,24}. The gel shown in **a** was blotted to nylon and the blot was stripped and reprobed to provide the data for **b**–**e**. **f**, Dnmt3L-independent methylation of the intergenic differentially methylated region of the *Dlk1-Gtl2*-imprinted cluster. **g**, Partial demethylation of the *H19* differentially methylated region in Dnmt3L-deficient male germ cells. The DNA blot hybridization at the left shows partial demethylation at *HpaII* sites between –766 and –3,994 bases 5' of the *H19* start site. The bisulphite genomic sequencing data at the right show partial demethylation of CpG sites within the region –3,486 to –3,908 base pairs 5' of the *H19* start site that contains a CTCF-binding site. A total of 39 bisulphite-converted DNA molecules were sequenced for the control and 45 for the Dnmt3L mutant. The 5' LTR IAP probe was as described¹³, and the LINE-1 5' untranslated-region probe was a PCR product that spanned the promoter region of a type A LINE-1 element from nucleotides 515–1,628 in GenBank accession no. M13002. The *Dlk1-Gtl2* DMR probe covered positions 82,879–83,729 in GenBank accession no. AG320506. The *H19* DMR probe spanned the 5' region from position –766 to –3,994 upstream of the *H19* transcription start site in GenBank accession no. M19619. The major satellite probe was from plasmid pMR196 (ref. 25), and minor satellite probe was an end-labelled oligonucleotide of sequence 5'-AACAGTGTATATCAATGAGTTACAATGAG-3'.

and their methylation status was equivalent in wild-type and mutant germ cells (Fig. 3d, e). This indicates the independent regulation of *de novo* methylation of dispersed and tandem repeated sequences. Figure 3f shows that the intergenic imprinting control region (the differentially methylated region (DMR)) of the *Dlk1-Gtl2*-imprinted cluster⁹ was not detectably demethylated in *Dnmt3L*-deficient germ cells, and there was only partial (about 50%) demethylation of the DMR of the *H19* gene^{10,11} (Fig. 3g). This is in contrast to the phenotype of *Dnmt3L*-deficient females, in which meiosis and the methylation of repeated sequences are normal and obvious methylation defects are limited to single-copy sequences associated with maternally imprinted genes³.

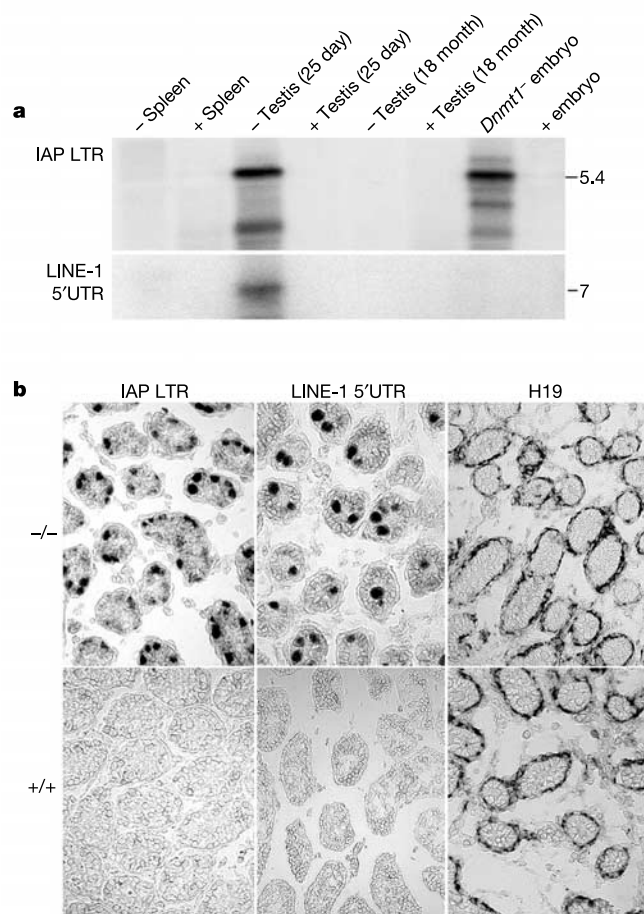


Figure 4 Transcriptional reactivation of retrotransposons in *Dnmt3L*-deficient germ cells. **a**, High-level expression of retrotransposons in *Dnmt3L*-deficient testis by RNA blot hybridization assay. Both LINE-1 and IAP transcripts are present at high concentrations in *Dnmt3L* mutant testis (third lane), whereas somatic cells of *Dnmt1*^{fl/fl} embryos at 9.5 d.p.c. (second lane from right) reactivate only IAP elements⁹. The predominant IAP subtype is the active $\Delta 1$ type at 5.4 kilobases (ref. 13). At 18 months, testes contain no detectable germ cells and there are no detectable retrotransposon transcripts (lane 5). **b**, Specific expression of LINE-1 and IAP transcripts in *Dnmt3L*-deficient germ cells. *In situ* hybridization against mutant (top row) and wild-type (bottom row) sections of testes from mice at 2 d.p.p. showed expression of high concentrations of LINE-1 and IAP transcripts in prospermatogonia. Further analysis showed expression of LINE-1 and IAP transcripts in dividing spermatogonia and spermatocytes as well. The imprinted *H19* gene was expressed at similar levels in mutant and control testes only in peritubular myoid cells, which confirms that the partial demethylation shown in Fig. 3g was not sufficient to reactivate *H19* expression in *Dnmt3L* mutant germ cells. Probes in both **a** and **b** were as in the DNA blot hybridization of Fig. 3.

The host defence hypothesis^{8,12} predicts that demethylation will reactivate retrotransposons in germ cells, as occurs in somatic cells whose genomes have become partly demethylated as a result of mutations in the *Dnmt1* gene¹³. RNA blot hybridization of RNA from whole testis showed that transcripts of IAP (reviewed in ref. 14) and LINE-1 (reviewed in ref. 15) retrotransposons were present in *Dnmt3L* mutant testes but undetectable in wild-type testes (Fig. 4a). Full-length transcripts of $\Delta 1$ IAP elements (5.4 kilobases) and A-type LINE-1 elements (about 7 kilobases), which are active retrotransposons, were present at high concentrations. *In situ* RNA hybridization revealed that IAP and LINE-1 retrotransposon transcripts were present at high concentrations only in mutant germ cells as early as 2 d.p.p., when the first spermatogonia begin to divide, and were present in all mutant spermatogonia and spermatocytes (Fig. 4b). Reactivation of retrotransposons in demethylated germ-cell genomes confirmed that cytosine methylation has an essential role in the suppression of retrotransposons in the mammalian germ line^{8,12}. Figure 4b also shows that the partial demethylation observed at the DMR of *H19* (Fig. 3) was not sufficient to reactivate this imprinted gene in germ cells, and expression of *H19* was confined to peritubular myoid cells in both mutant and control specimens. A more pronounced demethylation of *H19* in *Dnmt3L*-deficient male germ cells at earlier stages than those tested here was reported recently¹⁶.

The methylation defect in *Dnmt3L*-deficient germ cells is not due to a general defect in the *de novo* methylation of repeated sequences, as shown by the normal methylation of the tandem repeats of satellite DNA (which are largely free of retrotransposons); the defect is instead seen specifically at retrotransposons, which are interspersed repeated sequences. A major difference between populations of tandem and interspersed repeats is the abundance of homology–heterology boundaries in the latter; these boundaries have been proposed to be marks by which retrotransposons could be identified by the host¹⁷. Under this model, *Dnmt3L* is involved in a premeiotic genome scan that is proposed to occur in prospermatogonia within a brief perinatal period that establishes the lifelong silencing of retrotransposons. Silencing is proposed to involve heritable cytosine methylation mediated by the *Dnmt3L*-dependent recruitment and activation of *Dnmt3A* (refs 2, 18) and other silencing factors at homology–heterology boundaries. As the genome is scanned against itself, the frequency of contacts between dispersed repeats will follow first-order kinetics and will therefore increase as the second power of copy number. The premeiotic genome scan proposed here differs from a meiotic homology search in that the former identifies non-allelic repeats whereas the latter aligns homologous chromosomes. The postulated premeiotic genome scan is likely to occur without double-strand breaks or the invasion of duplex DNA by free DNA ends and is not therefore directly comparable to current models for homology search in meiosis. The most similar phenomena might be repeat-induced point mutation (RIP) in *Neurospora crassa*¹⁹, in which interactions between repeated sequences cause their mutual methylation and silencing. As in retrotransposon and imprinted gene methylation in mouse prospermatogonia, RIP occurs in the sexual cycle before the start of meiosis and is not known to share mechanisms with meiotic recombination.

Dnmt3L is required for normal male meiosis but is not expressed in spermatocytes, and the heritable defect in *Dnmt3L*-deficient male germ cells is likely to arise from a failure to establish methylation patterns on dispersed repeated sequences in perinatal prospermatogonia. The non-homologous synapsis that characterizes 1meiotic prophase in *Dnmt3L*-deficient spermatocytes could have several sources: it might result from a perturbation of gene expression, from illegitimate interactions between dispersed repeated sequences that were unmasked by demethylation²⁰, from single-strand or double-strand breaks produced during replicative retrotransposition or from an increased number of double-strand

breaks induced at ectopic transcription units²¹ activated by demethylation.

The data presented here show that Dnmt3L is required for heritable silencing of retrotransposons in male germ cells. Temporary deprivation of Dnmt3L through drug-induced conditional alleles of the *Dnmt3L* gene might allow the controlled mobilization of endogenous retrotransposons without meiotic catastrophe and could form the basis of a system of retrotransposon insertional mutagenesis that would be useful in forward genetic screens in the mouse. □

Received 22 May; accepted 26 July 2004; doi:10.1038/nature02886.
Published online 18 August 2004.

1. Aapola, U. *et al.* Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* **65**, 293–298 (2000).
2. Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of *de novo* DNA methyltransferases to establish maternal imprints in mice. *Development* **129**, 1983–1993 (2002).
3. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539 (2001).
4. La Salle, S. *et al.* Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. *Dev. Biol.* **268**, 403–415 (2004).
5. Dobson, M. J., Pearlman, R. E., Karaiskakis, A., Spyropoulos, B. & Moens, P. B. Synaptonemal complex proteins: occurrence, epitope mapping and chromosome disjunction. *J. Cell Sci.* **107**, 2749–2760 (1994).
6. Odorisio, T., Rodriguez, T. A., Evans, E. P., Clarke, A. R. & Burgoyne, P. S. Meiotic checkpoint monitoring synapsis eliminates spermatocytes via p53-independent apoptosis. *Nature Genet.* **18**, 257–261 (1998).
7. Enders, G. C. & May, J. J. Developmentally regulated expression of a mouse germ cell nuclear antigen examined from embryonic day 11 to adult in male and female mice. *Dev. Biol.* **163**, 331–340 (1994).
8. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
9. Lin, S. P. *et al.* Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dkl1-Gtl2* imprinted cluster on mouse chromosome 12. *Nature Genet.* **35**, 97–102 (2003).
10. Davis, T. L., Trasler, J. M., Moss, S. B., Yang, G. J. & Bartolomei, M. S. Acquisition of the *H19* methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* **58**, 18–28 (1999).
11. Ueda, T. *et al.* The paternal methylation imprint of the mouse *H19* locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells* **5**, 649–659 (2000).
12. Bestor, T. H. Cytosine methylation mediates sexual conflict. *Trends Genet.* **19**, 185–190 (2003).
13. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet.* **20**, 116–117 (1998).
14. Kuff, E. L. & Lueders, K. K. The intracisternal A-particle gene family: structure and functional aspects. *Adv. Cancer Res.* **51**, 183–276 (1988).
15. Ostertag, E. M. & Kazazian, H. H. Jr Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).
16. Kaneda, M. *et al.* Essential role for *de novo* methyltransferase 3a in paternal and maternal imprinting. *Nature* **429**, 900–903 (2004).
17. Bestor, T. H. & Tycko, B. Creation of genomic methylation patterns. *Nature Genet.* **12**, 363–367 (1996).
18. Chedin, F., Lieber, M. R. & Hsieh, C. L. The DNA methyltransferase-like protein DNMT3L stimulates *de novo* methylation by Dnmt3a. *Proc. Natl Acad. Sci. USA* **99**, 16916–16921 (2002).
19. Cambareri, E. B., Jensen, B. C., Schabach, E. & Selker, E. U. Repeat-induced G-C to A-T mutations in *Neurospora*. *Science* **244**, 1571–1575 (1989).
20. Maloel, L. & Rossignol, J.-L. Suppression of crossing-over by DNA methylation in *Ascomobolus*. *Genes Dev.* **12**, 1381–1389 (1998).
21. Petes, T. D. Meiotic recombination hot spots and cold spots. *Nature Rev. Genet.* **2**, 360–369 (2001).
22. Peters, A. H., Plug, A. W., van Vugt, M. J. & de Boer, P. A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. *Chromosome Res.* **5**, 66–68 (1997).
23. Ponzetto-Zimmerman, C. & Wolgemuth, D. J. Methylation of satellite sequences in mouse spermatogenic and somatic DNAs. *Nucleic Acids Res.* **12**, 2807–2822 (1984).
24. Sanford, J., Forrester, L. & Chapman, V. Methylation patterns of repetitive DNA sequences in germ cells of *Mus musculus*. *Nucleic Acids Res.* **12**, 2822–2835 (1984).
25. Pietras, D. F. *et al.* Construction of a small *Mus musculus* repetitive DNA library: identification of a new satellite sequence in *Mus musculus*. *Nucleic Acids Res.* **11**, 6965–6983 (1983).

Acknowledgements We thank B. Tycko for suggestions, P. Moens for antibodies against synaptonemal complex proteins and H1t and for discussions, G. Enders for antibodies against GCNA1, J. Walonoski for advice on *in situ* hybridization, and K. Anderson, M. Damelin and M. Goll for criticism of the manuscript. This work was supported by grants from the National Institutes of Health to T.H.B. and by a fellowship from the Rett Syndrome Research Foundation to D.B.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.H.B. (thb12@columbia.edu). *Dnmt3L^{tm1Bes}* mutant mice are available from the Jackson Labs (<http://jaxmice.jax.org/>).

Transcriptional regulatory code of a eukaryotic genome

Christopher T. Harbison^{1,2*}, D. Benjamin Gordon^{1*}, Tong Ihn Lee¹, Nicola J. Rinaldi^{1,2}, Kenzie D. Macisaac³, Timothy W. Danford³, Nancy M. Hannett¹, Jean-Bosco Tagne¹, David B. Reynolds¹, Jane Yoo¹, Ezra G. Jennings¹, Julia Zeitlinger¹, Dmitry K. Pokholok¹, Manolis Kellis^{1,3,4}, P. Alex Rolfe³, Ken T. Takusagawa³, Eric S. Lander^{1,2,4}, David K. Gifford^{3,4}, Ernest Fraenkel^{1,3} & Richard A. Young^{1,2,4}

¹Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

²Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

³MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, Massachusetts 02139, USA

⁴Broad Institute, One Kendall Square, Building 300, Cambridge, Massachusetts 02139, USA

* These authors contributed equally to this work

DNA-binding transcriptional regulators interpret the genome's regulatory code by binding to specific sequences to induce or repress gene expression¹. Comparative genomics has recently been used to identify potential *cis*-regulatory sequences within the yeast genome on the basis of phylogenetic conservation^{2–6}, but this information alone does not reveal if or when transcriptional regulators occupy these binding sites. We have constructed an initial map of yeast's transcriptional regulatory code by identifying the sequence elements that are bound by regulators under various conditions and that are conserved among *Saccharomyces* species. The organization of regulatory elements in promoters and the environment-dependent use of these elements by regulators are discussed. We find that environment-specific use of regulatory elements predicts mechanistic models for the function of a large population of yeast's transcriptional regulators.

We used genome-wide location analysis^{7–10} to determine the genomic occupancy of 203 DNA-binding transcriptional regulators in rich media conditions and, for 84 of these regulators, in at least 1 of 12 other environmental conditions (Supplementary Table 1, Supplementary Fig. 1; http://web.wi.mit.edu/young/regulatory_code). These 203 proteins are likely to include nearly all of the DNA-binding transcriptional regulators encoded in the yeast genome. Regulators were selected for profiling in an additional environment if they were essential for growth in that environment or if there was other evidence implicating them in the regulation of gene expression in that environment. The genome-wide location data identified 11,000 unique interactions between regulators and promoter regions at high confidence ($P \leq 0.001$).

To identify the *cis*-regulatory sequences that are likely to serve as recognition sites for transcriptional regulators, we merged information from genome-wide location data, phylogenetically conserved sequences, and prior knowledge (Fig. 1a). We used six motif discovery methods^{11–13} to discover 68,279 DNA sequence motifs for the 147 regulators that bound more than ten probes (Supplementary Methods, Supplementary Fig. 2). From these motifs we derived the most likely specificity for each regulator through clustering and stringent statistical tests. This motif discovery process identified highly significant ($P \leq 0.001$) motifs for each of 116 regulators. We determined a single high-confidence motif for 65 of these regulators by using additional criteria including the requirement for conservation across three of four related yeast species. Examples of discovered and rediscovered motifs are depicted in Fig. 1b, and comparisons of the discovered motifs with those described previously are shown in Supplementary Table 2.

The discovered motifs provide significantly more information than was previously available; for 21 of the regulators there was no prior specificity information in the literature, and detailed probability matrices had previously been determined for only 17 regulators for which we report motifs¹⁴. For Cin5, which showed the largest difference between the computationally derived motif (TTACRTAA) and the previously reported site (TTACTAA; Supplementary Table 2), we found that the motif we report is also the preferred target *in vitro* (Supplementary Fig. 3). We supplemented the discovered motifs with additional motifs from the literature that also passed conservation tests, and we used this compendium of sequence motifs for 102 regulators (Supplementary Table 3) in all subsequent analysis.

We constructed an initial version of the transcriptional regulatory code by mapping on the yeast genome sequence the motifs that are bound by regulators at high confidence ($P \leq 0.001$) and that are conserved among *sensu stricto* *Saccharomyces* species (Fig. 2; http://web.wi.mit.edu/fraenkel/regulatory_map). This map includes 3,353 interactions within 1,296 promoter regions. Maps of regulatory sites encompassing larger numbers of promoters, constructed with lower-confidence information, can also be viewed on the authors' website. Because the information used to construct the map includes binding data from multiple growth environments, the map describes transcriptional regulatory potential within the

genome. During growth in any one environment, only subsets of the binding sites identified in the map are occupied by transcriptional regulators, as we describe in more detail below.

Where the functions of specific transcriptional regulators were established previously, the functions of the genes they bind in the regulatory map are highly consistent with this prior information. For example, the amino-acid biosynthetic regulators Gcn4 and Leu3 bind to sites in the promoter of *BAP2* (chromosome II), which encodes an amino-acid transporter (Fig. 2a). Six well-studied cell cycle transcriptional regulators bind to the promoter for *YHP1* (chromosome IV), which has been implicated in the regulation of the G1 phase of the cell cycle. The regulator of respiration Hap5 binds upstream of *COX4* (chromosome VII), which encodes a component of the respiratory electron transport chain. Where regulators with established functions bind to genes of unknown function, these target genes are newly implicated in such functional processes.

The utility of combining regulator binding data and sequence conservation data is illustrated in Fig. 2b. All sequences matching the regulator DNA binding specificities described in this study (Supplementary Table 3) that occur within the 884-base-pair intergenic region upstream of the gene *BAP2* are shown in the upper panel. The subset of these sequences that have been conserved in multiple yeast species, and are therefore likely candidates for

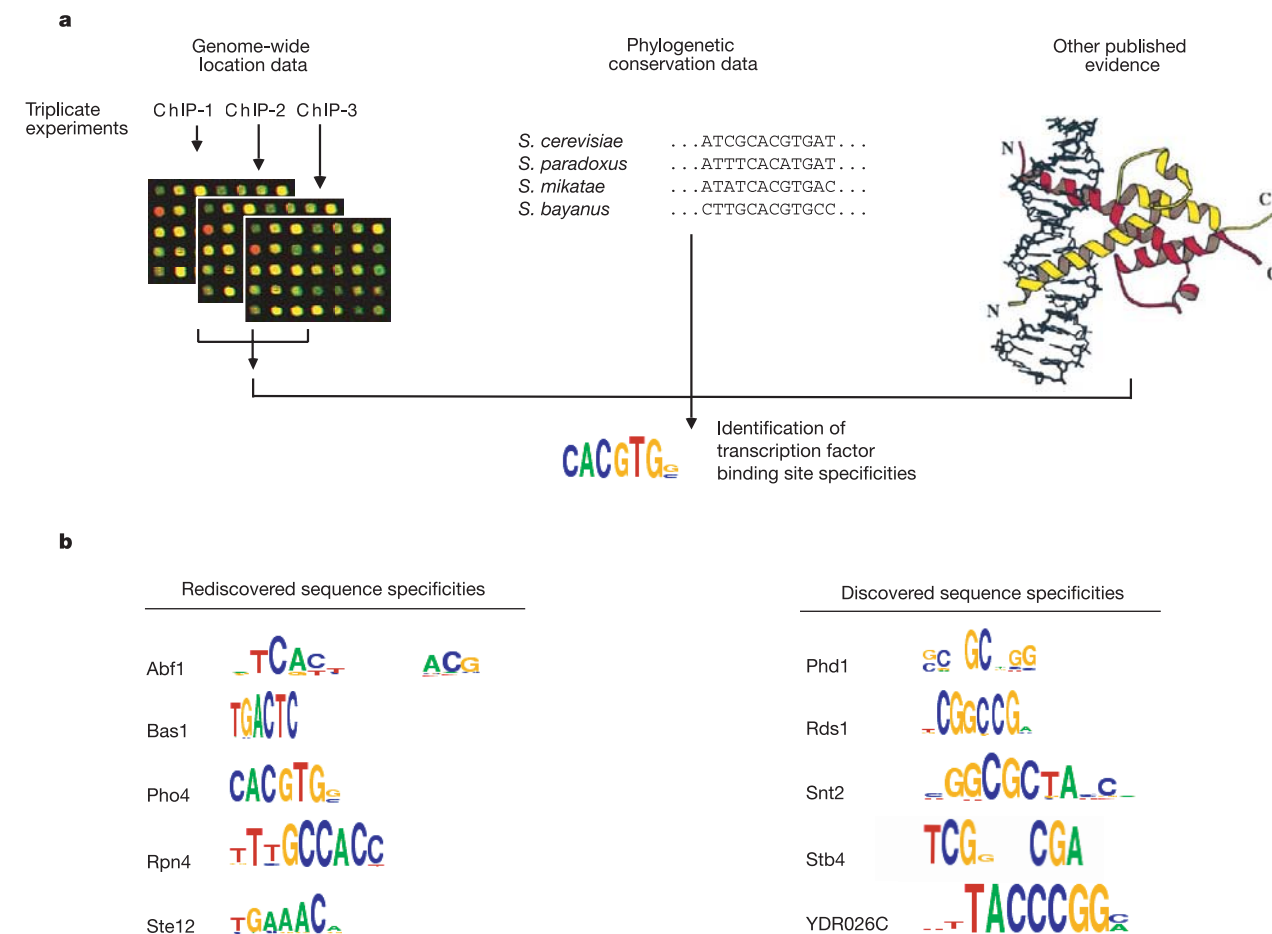


Figure 1 Discovering binding-site specificities for yeast transcriptional regulators. **a**, *Cis*-regulatory sequences likely to serve as recognition sites for transcriptional regulators were identified by combining information from genome-wide location data, phylogenetically conserved sequences and previously published evidence, as described in Supplementary Methods. The compendium of regulatory sequence motifs can be found in

Supplementary Table 3. **b**, Selected sequence specificities that were rediscovered and were newly discovered are shown. The total height of the column is proportional to the information content of the position, and the individual letters have a height proportional to the product of their frequency and the information content³⁰.

regulator interactions, is shown in the middle panel of Fig. 2b. The presence of these conserved regulatory sites indicates the potential for regulation through this sequence but does not indicate whether the site is actually bound by a regulator under some growth condition. The incorporation of binding information (Fig. 2b, bottom panel) identifies those conserved sequences that are used by regulators in cells grown under the conditions examined.

The distribution of binding sites for transcriptional regulators reveals constraints on the organization of these sites in yeast promoters (Fig. 2c). Binding sites are not uniformly distributed over the promoter regions but instead show a sharply peaked distribution. Very few sites are located in the region 100 base pairs (bp) upstream of protein-coding sequences. This region typically includes the transcription start site and is bound by the

transcription initiation apparatus. The vast majority (74%) of the transcriptional regulator binding sites lie between 100 and 500 bp upstream of the protein-coding sequence, far more than would be expected at random (53%). Regions further than 500 bp upstream contain fewer binding sites than would be expected at random. It seems that yeast transcriptional regulators function at short distances along the linear DNA, a property that reduces the potential for inappropriate activation of nearby genes.

We note that specific arrangements of DNA binding sites occur within promoters, and we suggest that these promoter architectures provide clues to regulatory mechanisms (Fig. 3). For example, the presence of a DNA binding site for a single regulator is the simplest promoter architecture and, as might be expected, we found that sets of genes with this feature are often involved in a common biological

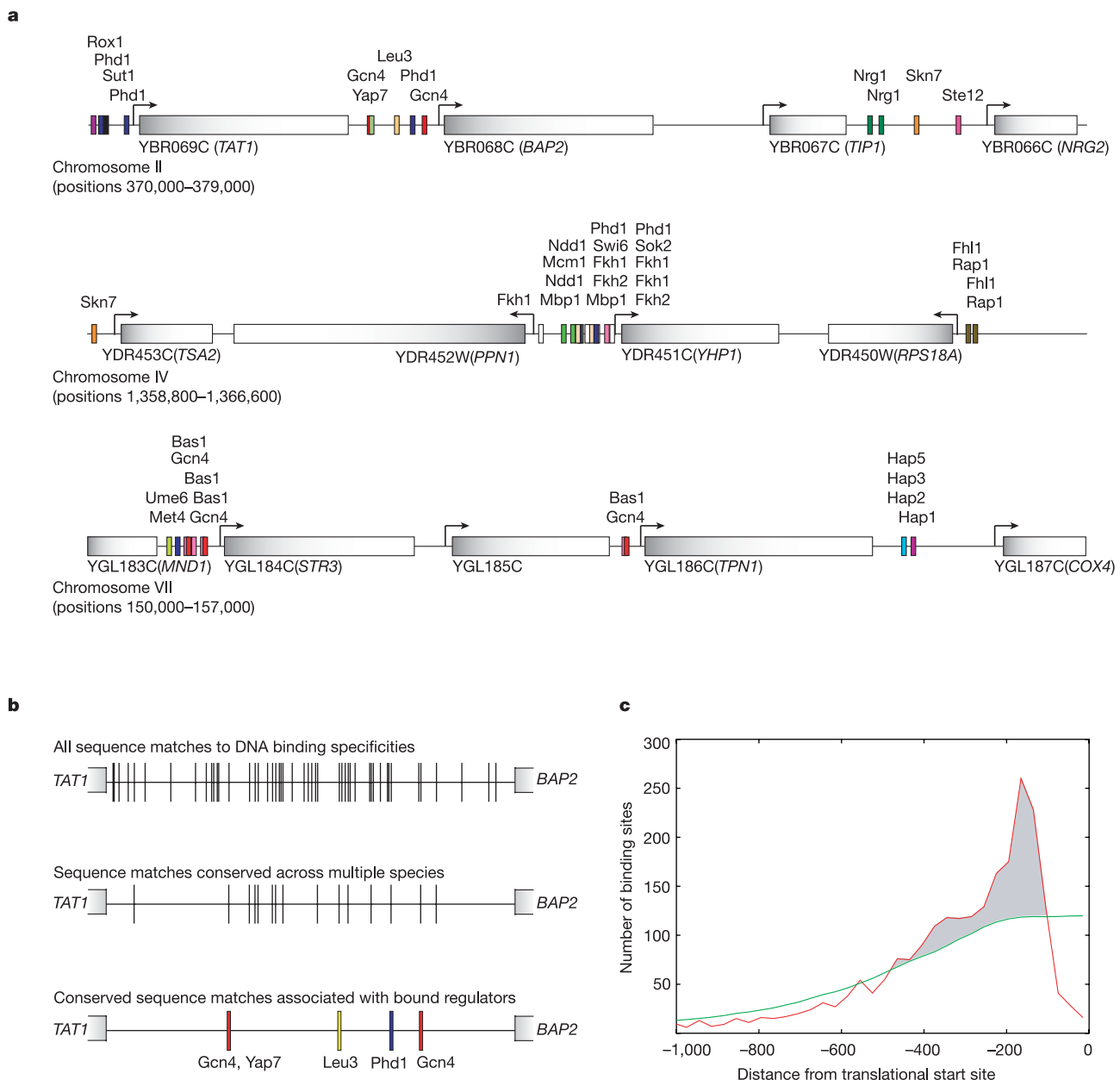


Figure 2 Drafting the yeast transcriptional regulatory map. **a**, Portions of chromosomes illustrating locations of genes (grey rectangles) and conserved DNA sequences (coloured boxes) bound *in vivo* by transcriptional regulators. **b**, Combining binding data and sequence conservation data. The diagram depicts all sequences matching a motif from our compendium (top), all such conserved sequences (middle) and all such conserved

sequences bound by a regulator (bottom). **c**, Regulator binding site distribution. The red line shows the distribution of distances from the start codon of open reading frames to binding sites in the adjacent upstream region. The green line represents a randomized distribution.

function (Supplementary Table 4). A second type of promoter architecture consists of repeats of a particular binding site sequence. Repeated binding sites have been shown to be necessary for stable binding by the regulator Dal80 (ref. 15). This repetitive promoter architecture can also permit a graded transcriptional response, as has been observed for the *HIS4* gene¹⁶. Several regulators, including Dig1, Mbp1 and Swi6, show a statistically significant preference for repetitive motifs (Supplementary Table 5). A third class of promoter contains binding sites for multiple different regulators. This promoter arrangement implies that the gene might be subject to combinatorial regulation, and we expect that in many cases the various regulators can be used to execute differential responses to varied growth conditions. Indeed, we note that many of the genes in this category encode products that are required for multiple metabolic pathways and are regulated in an environment-specific fashion. In the fourth type of promoter architecture we discuss here, binding sites for specific pairs of regulators occur more frequently within the same promoter regions than would be expected by chance (Supplementary Table 6). This 'co-occurring' motif architecture implies that the two regulators interact physically or have related functions at multiple genes.

By conducting genome-wide binding experiments for some regulators under multiple cell-growth conditions, we learned that regulator binding to a subset of the regulatory sequences is highly dependent on the environmental conditions of the cell (Supplementary Fig. 4). We observed four common patterns of regulator binding behaviour (Fig. 4, Supplementary Table 7). Prior information about the regulatory mechanisms employed by well-studied regulators in each of the four groups suggests hypotheses to account for the environment-dependent binding behaviour of the other regulators.

'Condition-invariant' regulators bind essentially the same set of promoters (within the limitations of noise) in two different growth environments (Fig. 4). Leu3, which is known to regulate genes involved in amino-acid biosynthesis, is among the best studied of the regulators in this group. Binding of Leu3 *in vivo* has been shown to be necessary but not sufficient for the activation of Leu3-regulated genes¹⁷. Rather, regulatory control of these genes requires the association of a leucine metabolic precursor with Leu3 to convert it from a negative to a positive regulator. We note that other zinc cluster type regulators that show 'condition-invariant' behaviour are known to be regulated in a similar manner^{18,19}. It is therefore reasonable to propose that the activation or repression functions of some of the other regulators in this class have requirements in addition to DNA binding.

'Condition-enabled' regulators do not bind the genome detectably under one condition, but bind a substantial number of promoters with a change in environment. Msn2 is among the best-studied regulators in this class, and the mechanisms involved in Msn2-dependent transcription provide clues to how the other regulators in that class might operate. Msn2 is excluded from the nucleus when cells grow in the absence of stresses but accumulates rapidly in the nucleus when cells are subjected to stress^{20,21}. This condition-enabled behaviour was also observed for the thiamine biosynthetic regulator Thi2, the nitrogen regulator Gat1 and the developmental regulator Rim101. We suggest that many of these transcriptional regulators are regulated by nuclear exclusion or by another mechanism that would cause this extreme version of condition-specific binding.

'Condition-expanded' regulators bind to a core set of target promoters under one condition but bind an expanded set of promoters under another condition. Gcn4 is the best-studied of the regulators that fall into this 'expanded' class. The levels of Gcn4 are reported to increase sixfold when yeast cells are introduced into media with limiting nutrients²², owing largely to increased nuclear protein stability^{21,23}, and under this condition we find that Gcn4 binds to an expanded set of genes. The probes bound when Gcn4

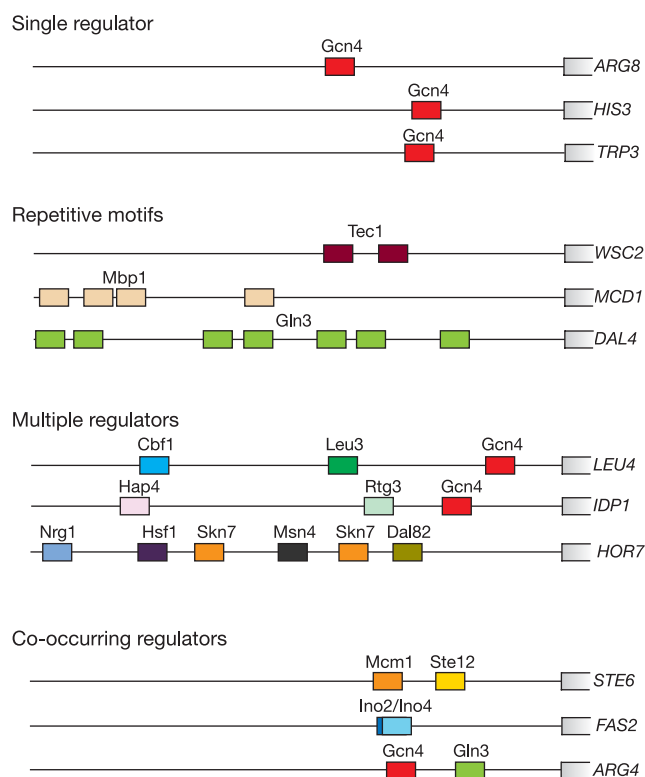


Figure 3 Yeast promoter architectures: single regulator architecture, promoter regions that contain one or more copies of the binding site sequence for a single regulator; repetitive motif architecture, promoter regions that contain multiple copies of a binding site sequence of a regulator; multiple regulator architecture, promoter regions that contain one or more copies of the binding site sequences for more than one regulator; co-occurring regulator architecture, promoters that contain binding site sequences for recurrent pairs of regulators. For the purposes of illustration, not all sites are shown and the scale is approximate. Additional information can be found in Supplementary Tables 4–6.

levels are low contain better matches to the known Gcn4-binding site than probes that are bound exclusively at higher protein concentrations, which is consistent with a simple model for specificity based on intrinsic protein affinity and protein concentration (Supplementary Fig. 5). The expansion of binding sites by many of the regulators in this class might reflect increased levels of the regulator available for DNA binding.

'Condition-altered' regulators exhibit an altered preference for the set of promoters bound in two different conditions. Ste12 is the best-studied of the regulators whose binding behaviour falls into this 'altered' class. Depending on the interactions with other regulators, the specificity of Ste12 can change and alter its cellular function²⁴. For example, under filamentous growth conditions, Ste12 interacts with Tec1, which has its own DNA-binding specificity²⁵. This condition-altered behaviour was also observed for the transcriptional regulators Aft2, Skn7 and Ume6. We propose that the binding specificity of many of the transcriptional regulators might be altered through interactions with other regulators or through modifications (such as chemical) that are dependent on environment.

Substantial portions of eukaryotic genome sequence are believed to be regulatory^{2,3,26}, but the DNA sequences that actually contribute to regulation of genome expression have been ill-defined. By mapping the DNA sequences bound by specific regulators in various environments, we identify the regulatory potential embedded in the genome and provide a framework for modelling the mechanisms

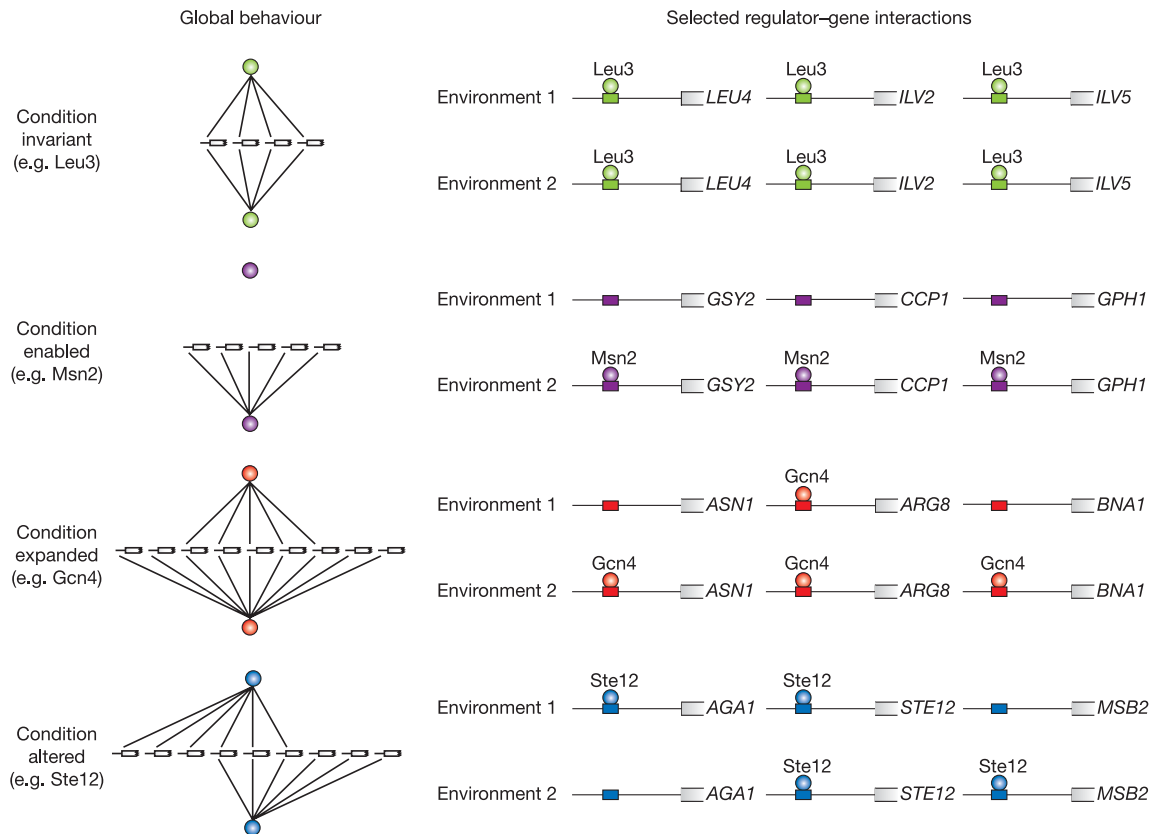


Figure 4 Environment-specific use of the transcriptional regulatory code. Four patterns of genome-wide binding behaviour are depicted on the left, where transcriptional regulators are represented by coloured circles and are placed above and below a set of target genes/promoters. The lines between the regulators and the target genes/promoters represent

binding events. Specific examples of the environment-dependent behaviours are depicted on the right. Coloured circles represent regulators and coloured boxes represent their DNA binding sequences within specific promoter regions. We note that regulators might exhibit different behaviours when different pairs of conditions are compared.

that contribute to global gene expression. We expect that the approaches used here to map regulatory sequences in yeast can also be used to map the sequences that control genome expression in higher eukaryotes. □

Methods

Strain information

For each of the 203 regulators, strains were generated in which a repeated Myc-epitope-coding sequence was integrated into the endogenous gene encoding the regulator. Polymerase chain reaction (PCR) constructs containing the Myc-epitope-coding sequence and a selectable marker flanked by regions of homology to either the 5' or 3' end of the targeted gene were transformed into the W303 yeast strain Z1256 (refs 8, 9). Genomic integration and expression of the epitope-tagged protein were confirmed by PCR and western blotting, respectively.

Genome-wide location analysis

Genome-wide location analysis was performed as described previously^{8,9}. Bound proteins were crosslinked by formaldehyde to DNA *in vivo*, followed by cell lysis and sonication to shear DNA. Crosslinked material was immunoprecipitated with an anti-Myc antibody, followed by reversal of the crosslinks to separate DNA from protein. Immunoprecipitated DNA and DNA from an unenriched sample were amplified and differentially fluorescently labelled by ligation-mediated PCR. These samples were hybridized to a microarray consisting of spotted PCR products representing the intergenic regions of the *Saccharomyces cerevisiae* genome. Relative intensities of spots were used as the basis for an error model that assigns a probability score (P) to binding interactions. All microarray data are available from ArrayExpress and from the authors' website (http://web.wi.mit.edu/young/regulatory_code).

Growth environments

We profiled all 203 regulators in rich medium. In addition, we profiled 84 regulators in at least one other environmental condition. The list of regulators is given in Supplementary Table 1.

Regulator binding specificity

The putative specificities of regulators were identified by applying a suite of motif

discovery programs to the intergenic sequences identified by the binding data. The resulting specificity predictions were filtered for significance with uniform metrics and then clustered to yield representative motifs (Supplementary Fig. 2).

We used six methods to identify the specific sequences bound by regulators: AlignACE¹¹, MEME¹³, MDscan¹², the method in ref. 2 and two additional new methods that incorporate conservation data: MEME_c and CONVERGE. MEME_c uses the existing MEME program without change but applies it to a modified set of sequences in which bases that are not conserved in the *sensu stricto* *Saccharomyces* species were replaced with the letter 'N'. CONVERGE is a novel expectation-maximization (EM)-based algorithm for discovering specificities by using sequence information from multiple genomes. Rather than searching for sites that are identical across the *sensu stricto* species, as occurs with MEME_c, CONVERGE searches for loci at which all aligned sequences are consistent with the same specificity model. See Supplementary Methods for runtime parameters and additional details for all of these methods.

Each of the programs we used attempts to measure the significance of its results with one or more statistical scores. However, we observed that these programs report results with high scores even when applied to random selections of intergenic regions. To distinguish the true motifs we chose a set of statistical measures that are described in Supplementary Methods and converted these scores into the empirical probability that a motif with a similar score could be found by the same program in randomly selected sequences. To estimate these P values we ran each program 50 times on randomly selected sets of sequences of various sizes. We accepted only those motifs that were judged to be significant by these scores ($P \leq 0.001$).

Significant motifs from all programs were pooled and clustered with the use of a k -medoids algorithm. Aligned motifs within each cluster were averaged to produce consensus motifs and filtered according to their conservation. This procedure typically produced several distinct consensus motifs for each regulator. To choose a single specificity for each regulator we compared the results with information in the TRANSFAC²⁷, YPD²⁸ and SCPD²⁹ databases. When no prior information was available we chose the specificity with the most significant statistical score.

Regulatory code

Potential binding sites were included in the map of the regulatory code if they satisfied two criteria. First, a locus had to match the specificity model for a regulator in the *S. cerevisiae* genome and at least two other *sensu stricto* *Saccharomyces* genomes with a score of at least 60% of the maximum possible. Second, the locus had to lie in an intergenic region that also contained a probe bound by the corresponding regulator in any condition ($P \leq 0.001$). All

analyses of promoter architecture and environment-specific binding were based on this map and can be found in Supplementary Information.

Received 11 March; accepted 1 July 2004; doi:10.1038/nature02800.

- Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
- Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76 (2003).
- Pritsker, M., Liu, Y. C., Beer, M. A. & Tavazoie, S. Whole-genome discovery of transcription factor binding sites by network-level conservation. *Genome Res.* **14**, 99–108 (2004).
- Wang, T. & Stormo, G. D. Combining phylogenetic data with co-regulated genes to identify regulatory motifs. *Bioinformatics* **19**, 2369–2380 (2003).
- Blanchette, M. & Tompa, M. FootPrinter: A program designed for phylogenetic footprinting. *Nucleic Acids Res.* **31**, 3840–3842 (2003).
- Iyer, V. R. *et al.* Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538 (2001).
- Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
- Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
- Lieb, J. D., Liu, X., Botstein, D. & Brown, P. O. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nature Genet.* **28**, 327–334 (2001).
- Roth, F. P., Hughes, J. D., Estep, P. W. & Church, G. M. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nature Biotechnol.* **16**, 939–945 (1998).
- Liu, X. S., Brutlag, D. L. & Liu, J. S. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nature Biotechnol.* **20**, 835–839 (2002).
- Bailey, T. L. & Elkan, C. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **Vol. 3** 21–29 (AAAI Press, Menlo Park, California, 1995).
- Knuppel, R., Dietze, P., Lehnberg, W., Frech, K. & Wingender, E. TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J. Comput. Biol.* **1**, 191–198 (1994).
- Cunningham, T. S. & Cooper, T. G. The *Saccharomyces cerevisiae* DAL80 repressor protein binds to multiple copies of GATAA-containing sequences (URSGATA). *J. Bacteriol.* **175**, 5851–5861 (1993).
- Donahue, T. E., Daves, R. S., Lucchini, G. & Fink, G. R. A short nucleotide sequence required for regulation of HIS4 by the general control system of yeast. *Cell* **32**, 89–98 (1983).
- Kirkpatrick, C. R. & Schimmel, P. Detection of leucine-independent DNA site occupancy of the yeast Leu3p transcriptional activator *in vivo*. *Mol. Cell. Biol.* **15**, 4021–4030 (1995).
- Axelrod, J. D., Majors, J. & Brandriss, M. C. Proline-independent binding of PUT3 transcriptional activator protein detected by footprinting *in vivo*. *Mol. Cell. Biol.* **11**, 564–567 (1991).
- Ma, J. & Ptashne, M. The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80. *Cell* **50**, 137–142 (1987).
- Beck, T. & Hall, M. N. The TOR signalling pathway controls nuclear localization of nutrient-regulated transcription factors. *Nature* **402**, 689–692 (1999).
- Chi, Y. *et al.* Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. *Genes Dev.* **15**, 1078–1092 (2001).
- Albrecht, G., Mosch, H. U., Hoffmann, B., Reusser, U. & Braus, G. H. Monitoring the Gcn4 protein-mediated response in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**, 12696–12702 (1998).
- Kornitzer, D., Raboy, B., Kulka, R. G. & Fink, G. R. Regulated degradation of the transcription factor Gcn4. *EMBO J.* **13**, 6021–6030 (1994).
- Zeitlinger, J. *et al.* Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* **113**, 395–404 (2003).
- Baur, M., Esch, R. K. & Errede, B. Cooperative binding interactions required for function of the Tyl1 sterile responsive element. *Mol. Cell. Biol.* **17**, 4330–4337 (1997).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
- Hodges, P. E., McKee, A. H., Davis, B. P., Payne, W. E. & Garrels, J. I. The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data. *Nucleic Acids Res.* **27**, 69–73 (1999).
- Zhu, J. & Zhang, M. Q. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics* **15**, 607–611 (1999).
- Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Ideker and S. McCuine for help in selecting regulators to study in environmental conditions; E. Herbolzheimer, G. Bell, R. Latek and F. Lewitter for computational assistance; and E. McReynolds for technical assistance. E.F. is a Whitehead Fellow and was funded in part by Pfizer. D.B.G. was supported by a NIH/NIGMS NRSA award. This work was supported by an NIH grant.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to R.A.Y. (young@wi.mit.edu) or E.F. (efraenkel@wi.mit.edu). ArrayExpress number E-WMIT-10 has been given for microarray data.

Structure of the acrosomal bundle

Michael F. Schmid¹, Michael B. Sherman^{1*}, Paul Matsudaira² & Wah Chiu¹

¹National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA

²Whitehead Institute, Department of Biology and Division of Biological Engineering MIT, Cambridge, Massachusetts 02142, USA

* Present address: Purdue University, Department of Biological Sciences, West Lafayette, Indiana 47907-139, USA

In the unactivated *Limulus* sperm, a 60- μ m-long bundle of actin filaments crosslinked by the protein scruin is bent and twisted into a coil around the base of the nucleus. At fertilization, the bundle uncoils and fully extends in five seconds to support a finger of membrane known as the acrosomal process. This biological spring is powered by stored elastic energy and does not require the action of motor proteins or actin polymerization¹. In a 9.5-Å electron cryomicroscopic structure of the extended bundle, we show that twist, tilt and rotation of actin-scrutin subunits deviate widely from a ‘standard’ F-actin filament. This variability in structural organization allows filaments to pack into a highly ordered and rigid bundle in the extended state and suggests a mechanism for storing and releasing energy between coiled and extended states without disassembly.

The actin bundle of the *Limulus* acrosomal process is an unusual actin-based engine that is highly ordered in its structure². The coiled state has a difference in average twist of 0.23° per subunit relative to the extended form. Change in the twist accompanies uncoiling of the bundle into the final extended state³. Mechanical measurements of the extended bundle yield a Young’s modulus (a measure of stiffness) of 2 GPa (refs 4–6). This stiffness enables the bundle to penetrate the jelly layer surrounding the egg membrane. A 40-Å tomographic reconstruction⁷ revealed periodic and dense packing of the scruin-crosslinked actin filaments. However, to understand all these functions, a higher resolution structure is necessary.

An electron micrograph of the native extended bundle shows its crystallinity (Fig. 1a). We merged data from 153 such bundles at various orientations⁸ to produce a three-dimensional map at a resolution of 9.5 Å (Fig. 1b). As far as we know, this is the highest resolution of any electron cryomicroscopic reconstruction containing an F-actin filament. We can identify each actin subunit and both domains of scruin, referred to as elongated (E) and spheroidal (S) on the basis of their shape⁹, and secondary structure elements of both molecules.

We used correlation analysis and non-crystallographic symmetry (NCS) averaging within the asymmetric unit to determine the azimuthal and axial positions of the 14 actin-scrutin subunits. The azimuthal rotation between adjacent subunits ranges from –176.5° to –142.5° (Fig. 1c, d), and thus deviates from the canonical actin helix by –10.4° under-twist to +23.6° over-twist. Nevertheless, the rise per subunit in the axial direction had less than one pixel (1.33 Å) deviation. This amount of individual azimuthal distortion is larger than the distortions found in averaged reconstructions from F-actin filaments^{10–15}, but such studies cannot reveal individual subunit variations.

On the basis of this two-dimensional cross-correlation, we created an average actin monomer from the 14 subunits in the asymmetric unit. In this map (Fig. 2a), three α -helices in subdomain 1 of actin were detected by an α -helix search (using the program ‘helixhunter’)¹⁶. Their locations match α -helices in the actin crystal structure (amino acids 79–91, 113–125 and 137–144)¹⁷. Because we obtained this map without referring to the actin crystal structure during either the reconstruction process⁸ or the NCS

averaging, finding these α -helices in locations expected for an authentic actin molecule (Supplementary Information) validated our analysis and the reconstruction.

We subsequently aligned and averaged the actin more accurately with a six-dimensional rigid body search of the 14 actin subunits, using the atomic coordinates¹⁸. In this improved map we were able to identify three additional α -helices (182–194, 203–216 and 367–373) (Fig. 2b). Furthermore, the side view (Fig. 2c) shows a density near residues 264–273 projecting from the back of the subunit. This protrusion corresponds to the hypothesized “hydrophobic plug” that was built in an F-actin model on the basis of fibre diffraction¹⁸. In all actin X-ray crystal structures the “plug” region has other conformations because the interstrand contacts of an F-actin filament are non-existent in a crystal. In our reconstruction, the plug is positioned to stabilize the two strands of the actin filament.

Finally, we ran the program ‘foldhunter’ again, searching the averaged map (Fig. 2b, c) with pairs of subdomains (subdomains 1,

2 and 3, 4). The coordinates derived from this search (Fig. 2d) reveal that in the average structure, subdomains 2 and 4 at the top of the molecule are brought 2–3 Å closer together by a rotation of 3° of subdomains 3 and 4, relative to subdomains 1 and 2. This ‘closed’ structure is characteristic of nucleotide-bound actin conformation^{19–21} and has been found in actomyosin filaments²². Intriguingly, actin subdomain movement in muscle in response to myosin binding has recently been inferred from fluorescence depolarization experiments²³. Here it is probably also influenced by scruin interactions. Biochemical evidence shows that ADP is the nucleotide present in the bundle⁵.

The orientations determined by the six-dimensional cross-correlation search reveal significant tilt and rotation of the actin subunits (Fig. 3 and Supplementary Movie). Compared with a standard filament (Fig. 3a), actin subunits in the bundle (Fig. 3b) vary either in tilt (in and out of the plane of the actin molecule about a horizontal axis), which predominantly occurs in the yellow strand in Fig. 3, or in rotation (about the centre of the molecule), which is more common in the cyan strand. The average deviation from the orientation of the Holmes F-actin model¹⁸ in tilt or rotation was 11.3°. However, the hydrophobic plug (Fig. 2c) can still maintain contact with the opposite strand. Rotation of up to 12° was observed in averaged reconstructions of actin with actin depolymerizing factor¹⁴.

Actin’s ability to adopt different conformations by subdomain movements or gross changes of the F-actin helix is well-documented in crystals and in helically averaged filaments^{21,24,25}. We show how a ‘quasi-helical’ filament packs into a naturally occurring crystalline bundle. The two-stranded filament adjusts itself to distortions generated in the biological context of the acrosomal bundle. As predicted²⁶, actin has “random variable twist”, in that the azimuth in one subunit has no apparent relationship to that of the preceding or following subunits along the filament. In the bundle, the twist is

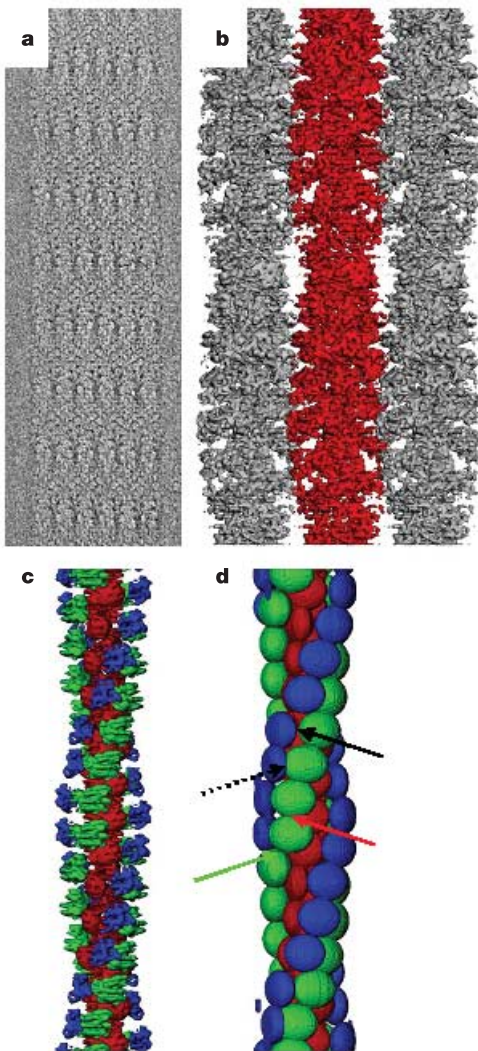


Figure 1 Acrosomal bundle reconstruction. **a**, Image of bundle near a major crystallographic zone. **b**, Three filaments from the crystal. Each unit cell (one scruin–actin filament, $146 \times 146 \times 765$ Å, $\gamma = 120^\circ$) contains 4.7 MDa. **c**, NCS averaged actin (red), scruin E (green) and scruin S domain (blue) placed back into one filament. **d**, Ellipses mark scruin E and S domains and actin in the above colours. The green arrow indicates overtwist and the red arrow indicates undertwist of E domains, which appear to be correlated between the scruin and actin. The black arrow shows larger E/S scruin spacing, and the dotted arrow shows smaller spacing.

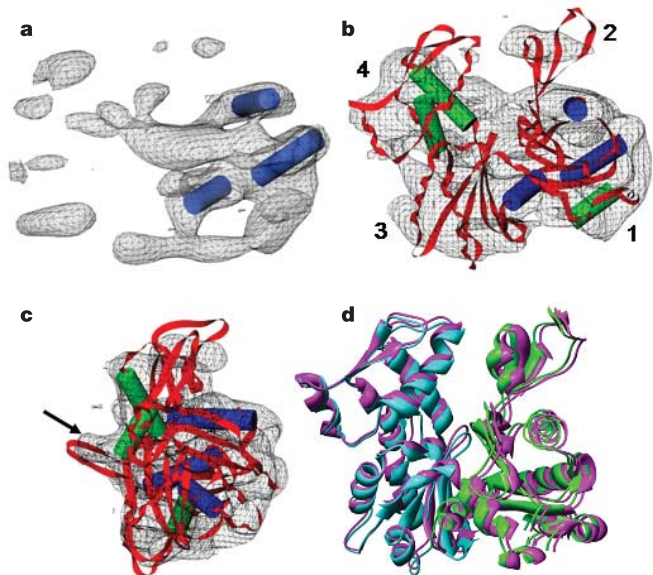


Figure 2 Averaged actin. **a**, Two-dimensional cross-correlation averaging. Putative helices are shown in blue. **b**, After a six-dimensional foldhunter¹⁶ search. Subdomains are labelled¹⁷. Holmes coordinates¹⁸ are shown in red ribbon. The three helices found in **a** are shown in blue, and those newly recognized in this map are green. Near-vertical helices are not resolved owing to higher B-factor (faster falloff in structure factor intensity) in the horizontal plane⁸. **c**, Side view showing the hydrophobic plug (arrow). **d**, Crystal structure (1ATN in purple)¹⁷ and model after subdomain fitting (cyan and green) to the NCS averaged map (**b**). Compared with the Holmes crystal structure¹⁷, actin in the acrosomal filament has a closed conformation.

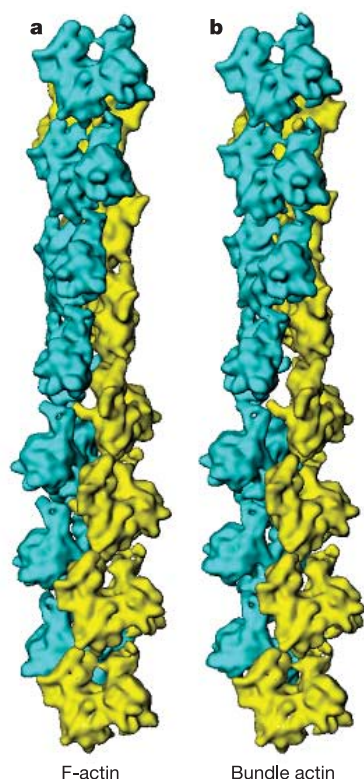


Figure 3 Actin filament distortion. **a**, **b**, The F-actin model of Holmes based on fibre diffraction (**a**) depicts the standard 13-unit, 6-turn helix of actin subunits. Two strands are coloured differently (yellow and cyan). From the results of the six-dimensional rigid body foldhunter search (**b**), actin subunits deviate significantly in position and in orientation from the canonical F-actin model. Subunits are predominantly tilted in the yellow strand, and rotated in the cyan strand (view movie in Supplementary Information).

indeed variable within an asymmetric unit but, conversely, the same pattern of variation repeats itself in the crystalline bundle. Furthermore, the variations (Fig. 3b) of the subunits appear to be coordinated along a strand, and asymmetric between strands (that is, one by tilt of its actin subunits, and the other by rotation of its subunits). As far as we know, this is the first direct evidence for symmetry-breaking between the two strands in an actin filament. The allowable actin distortion (Fig. 3b) affects the interactions and dynamics available to the bundle in its coiled and extended states.

Although the scriuin atomic coordinates are not yet available, we carried out a two-dimensional search with the scriuin domain densities using iterative cross-correlation and NCS averaging, as described above for the actin. Scriuin follows a distortion pattern similar to that of actin, with variable azimuth for both the S and E domains of each subunit (Fig. 1d). The averaged map of each domain (Fig. 4a, b) shows features that are consistent with a beta-propeller based structure homologous to kelch repeats²⁷. Confirmation of this domain structure requires a higher resolution map.

Placing the averaged actin and scriuin at the positions determined by the correlation results reveals the relationship among the molecules within an asymmetric unit. This averaged filament (Fig. 1c) shows that each scriuin domain interacts with actin subdomains within a filament at different radii, and perhaps with different functions, as previously observed and suggested^{9,28}. The interfilament contacts are primarily between the S domains of different filaments (Supplementary Information). This conclusion differs from a much lower resolution (~ 40 Å) reconstruction⁷. To a first approximation, the overtwist and undertwist patterns of the actin and scriuin in a filament are similar (Fig. 1d), suggesting that

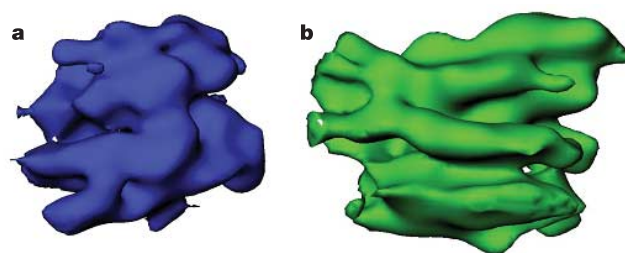


Figure 4 Averaged scriuin. **a**, **b**, Scriuin S (**a**) and E (**b**) domains derived by cross-correlation and NCS averaging. Each domain (50 and 60 kDa, respectively) is consistent with a beta-propeller motif (for example, galactose oxidase central domain-1GOF) at an equivalent resolution.

the scriuin packing causes the actin distortions.

The bundle is both stiff and conformationally variable. The Young's modulus of the extended bundle⁴⁻⁶ is like that of polyethylene or nylon. This stiffness combined with a force of nano-Newton scale provides the pressure to penetrate the jelly coat of the egg. Biochemical evidence suggests that calcium is the switch that changes scriuin structure or interactions, and triggers straightening of the bundle²⁹. Our model suggests that scriuin packing dictates actin filament structure and the organization of those filaments into the rigid bundle.

Figure 1d shows that the separation between the S and E domains in each scriuin molecule varies among different subunits in the asymmetric unit. Thus the S domain interacts in a variable way with the E domain and the actin of its own filament, and with the S domains from neighbouring filaments (Supplementary Information). Such variable interactions may be critical for the structures of both coiled and extended forms of the bundle and especially the transition between them.

The structure of the extended form of the bundle suggests how the coiled form can store energy. Distortion of actin is balanced against and coupled to the maintenance of interfilament scriuin contacts in both the coiled and the extended form. We predict that the actin will respond to the changes in scriuin during the acrosomal reaction and that transitions must occur along an energetically reasonable path. The acrosomal bundle was proposed to be a biological spring³⁰. Our studies have shown that it stores 6×10^{-13} joules of elastic energy. In the coiled state, the subunits are on average over-twisted with respect to the extended state, but under-twisted with respect to standard F-actin. The observed actin distortion (Fig. 3b) suggests that scriuin-decorated actin filaments in the bundle can store entropic or enthalpic energy in the actin helix. The extensive interactions between filaments could both 'lock-in' the energy and mediate the transition between coiled and extended conformations. We expect that conformational variability of the actin filament at the subunit-by-subunit level will be observed in other actin-containing machines, for example the muscle sarcomere, probably modulated and regulated by other components of the thin filament such as tropomyosin. □

Received 13 May; accepted 20 July 2004; doi:10.1038/nature02881.

1. Tilney, L. G. Actin filaments in the acrosomal reaction of *Limulus* sperm. *J. Cell Biol.* **64**, 289–310 (1975).
2. Schmid, M. F., Jakana, J., Matsudaira, P. & Chiu, W. Imaging frozen, hydrated acrosomal bundle from *Limulus* sperm at 7 Å resolution with a 400 kV electron cryomicroscope. *J. Mol. Biol.* **230**, 384–386 (1993).
3. DeRosier, D., Tilney, L. & Flicker, P. A change in the twist of the actin-containing filaments occurs during the extension of the acrosomal process in *Limulus* sperm. *J. Mol. Biol.* **137**, 375–389 (1980).
4. Shin, J. H., Mahadevan, L., So, P. T. & Matsudaira, P. Bending stiffness of a crystalline actin bundle. *J. Mol. Biol.* **337**, 255–261 (2004).
5. Shin, J. H., Mahadevan, L., Waller, G. S., Langsetmo, K. & Matsudaira, P. Stored elastic energy powers the 60-microm extension of the *Limulus polyphemus* sperm actin bundle. *J. Cell Biol.* **162**, 1183–1188 (2003).

6. Gardel, M. L. *et al.* Elastic behavior of cross-linked and bundled actin networks. *Science* **304**, 1301–1305 (2004).
7. Sherman, M. B. *et al.* The three-dimensional structure of the *Limulus* acrosomal process: a dynamic actin bundle. *J. Mol. Biol.* **294**, 139–149 (1999).
8. Schmid, M. F. Cross-correlation and merging of crystallographic reflections derived from cryoelectron micrographs of 3D crystals: application to the *Limulus* acrosomal bundle. *J. Struct. Biol.* **144**, 195–208 (2003).
9. Schmid, M. F., Agris, J. M., Jakana, J., Matsudaira, P. & Chiu, W. Three-dimensional structure of a single filament in the *Limulus* acrosomal bundle: scruin binds to homologous helix–loop–beta motifs in actin. *J. Cell Biol.* **124**, 341–350 (1994).
10. Galkin, V. E. *et al.* The location of ubiquitin in *Lethocerus* arthrin. *J. Mol. Biol.* **325**, 623–628 (2003).
11. Galkin, V. E. *et al.* The bacterial protein SipA polymerizes G-actin and mimics muscle nebulin. *Nature Struct. Biol.* **9**, 518–521 (2002).
12. Galkin, V. E. *et al.* The utrophin actin-binding domain binds F-actin in two different modes: implications for the spectrin superfamily of proteins. *J. Cell Biol.* **157**, 243–251 (2002).
13. Orlova, A. *et al.* Probing the structure of F-actin: cross-links constrain atomic models and modify actin dynamics. *J. Mol. Biol.* **312**, 95–106 (2001).
14. Galkin, V. E., Orlova, A., Lukyanova, N., Wriggers, W. & Egelman, E. H. Actin depolymerizing factor stabilizes an existing state of F-actin and can change the tilt of F-actin subunits. *J. Cell Biol.* **153**, 75–86 (2001).
15. McGough, A., Pope, B., Chiu, W. & Weeds, A. Cofilin changes the twist of F-actin: implications for actin filament dynamics and cellular function. *J. Cell Biol.* **138**, 771–781 (1997).
16. Jiang, W., Baker, M. L., Ludtke, S. J. & Chiu, W. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J. Mol. Biol.* **308**, 1033–1044 (2001).
17. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. Atomic structure of the actin:DNase I complex. *Nature* **347**, 37–44 (1990).
18. Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. Atomic model of the actin filament. *Nature* **347**, 44–49 (1990).
19. Dominguez, R. & Graceffa, P. Solution properties of TMR-actin: when biochemical and crystal data agree. *Biophys. J.* **85**, 2073–2074 (2003).
20. Graceffa, P. & Dominguez, R. Crystal structure of monomeric actin in the ATP state. Structural basis of nucleotide-dependent actin dynamics. *J. Biol. Chem.* **278**, 34172–34180 (2003).
21. Sablin, E. P. *et al.* How does ATP hydrolysis control actin's associations? *Proc. Natl Acad. Sci. USA* **99**, 10945–10947 (2002).
22. Holmes, K. C., Angert, I., Kull, F. J., Jahn, W. & Schroder, R. R. Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature* **425**, 423–427 (2003).
23. Borovikov, Y. S. *et al.* Fluorescence depolarization of actin filaments in reconstructed myofibers: the effect of S1 or pPDM-S1 on movements of distinct areas of actin. *Biophys. J.* **86**, 3020–3029 (2004).
24. Otterbein, L. R., Graceffa, P. & Dominguez, R. The crystal structure of uncomplexed actin in the ADP state. *Science* **293**, 708–711 (2001).
25. Egelman, E. H. Actin allostery again? *Nature Struct. Biol.* **8**, 735–736 (2001).
26. Egelman, E. H., Francis, N. & DeRosier, D. J. F-actin is a helix with a random variable twist. *Nature* **298**, 131–135 (1982).
27. Way, M., Sanders, M., Garcia, C., Sakai, J. & Matsudaira, P. Sequence and domain organization of scruin, an actin-cross-linking protein in the acrosomal process of *Limulus* sperm. *J. Cell Biol.* **128**, 51–60 (1995).
28. Bullitt, E. S., DeRosier, D. J., Coluccio, L. M. & Tilney, L. G. Three-dimensional reconstruction of an actin bundle. *J. Cell Biol.* **107**, 597–611 (1988).
29. Sanders, M., Way, M., Sakai, J. & Matsudaira, P. Characterization of the actin crosslinking properties of the scruin-calmodulin complex from the acrosomal process of *Limulus* sperm. *J. Biol. Chem.* **271**, 2651–2657 (1996).
30. Mahadevan, L. & Matsudaira, P. Motility powered by supramolecular springs and ratchets. *Science* **288**, 95–100 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This research is supported by the NCCR and NIGMS of NIH. We thank M. Baker for assistance in the helixhunter and foldhunter searches, and M. Dougherty for advice on graphical display.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.C. (wah@bcm.tmc.edu). The accession number is EMD-1088.

addendum

Pathways towards and away from Alzheimer's disease

Mark P. Mattson

Nature **430**, 631–639 (2004).

In the list of treatments given in Fig. 3 of this Review Article, I would like to add “Glutamate-receptor modulating agents (**2, 3, 4, 6**)”, where the numbers in bold indicate the sites of action in the pathogenesis pathway. □

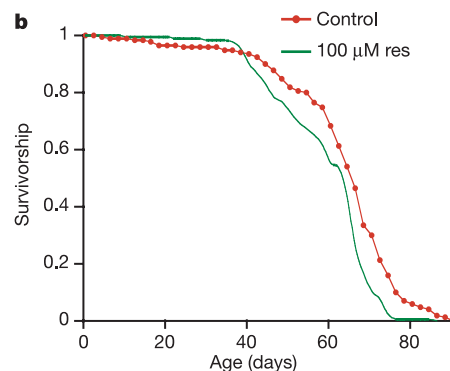
corrigendum

Sirtuin activators mimic caloric restriction and delay ageing in metazoans

Jason G. Wood, Blanka Rogina, Siva Lavu, Konrad Howitz, Stephen L. Helfand, Marc Tatar & David Sinclair

Nature **430**, 686–689 (2004).

There are errors in Fig. 4 of this Letter: panels a and d are correct; however, panel c was incorrectly published as a duplicate of panel a, and panel b should have been labelled as panel c instead. The new panel b is shown here. □



Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayouni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

The skills hunt

Earlier this summer, the New York Academy of Sciences and *Naturejobs* brought together potential employers and employees at a symposium called 'Where the jobs are'. During the various question-and-answer sessions, several young scientists asked separate panels representing academia, biotechnology and drug discovery how they could guarantee a clear career path for themselves. Although the questioners, panellists and disciplines were different at each session, the answers were surprisingly similar: do the best science you can, identify the key 'soft' or non-scientific skills associated with your ideal career path, and then find ways to obtain them.

But that's not the whole picture. Both employers and their prospective employees also have more immediate responsibilities to discharge. Students and postdocs, for example, first need to learn more about what their options are. Several of the 500-strong audience did not seem to know what is entailed by many of the leading non-bench scientific careers, let alone what skills they would need to do such jobs.

And several of the panellists noted that students often have difficulty telling the difference between drug discovery and development, or between business development and regulatory affairs. They admitted that each of their respective sectors could and should try harder to articulate just what these jobs entail and what skills and training young scientists need to do them.

As for stability and security, the panellists all agreed that hard work and the ability to adapt to change helped them to secure their current posts. Understanding specific trends can provide a starting place for the current crop of students and postdocs (see www.nyas.org), but a greater knowledge about what skills are lacking beyond academia or industry will move them closer to their career goals.

Paul Smaglik
Naturejobs editor



WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS